THE UNIVERSITY
of EDINBURGH

Characterising the *in vitro* conversion of mouse fibroblasts into induced neurons using a CRISPR/*Cas9* knock-out screen

Benedetta Carbone

Thesis presented for the degree of Doctor of Philosophy

MRC Centre for Regenerative Medicine

University of Edinburgh, 2019

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Benedetta Carbone

Edinburgh, August 2019

Acknowledgments

Huge thanks to my supervisor Kei for his help through this project, for pushing me to think critically and work to always improve. Thanks also to the Kaji lab for making these past 4 years a great experience. Thanks to all the previous lab members Nic, Tyson, Miguel and Eleni who welcomed me with enthusiasm and kindness. I'd also like to thank Suling, Spring, Lynsey, James, Linh, Anna, Haln and all the other members of the lab or of the lively break-out space who through the years took the time to share a cup of coffee, some words of encouragement and some laughs with me.

Thanks to my second supervisor Dr. Tilo Kunath and to my committee members Dr. Anna Williams and Dr. Steve Pollard for their scientific advice and support with the project. Thanks also to the everyone else within SCRM, the TC ladies and the Flow team for helping me through the ups and downs of managing my project. And thanks to the admin team, for their practical support and for their maintenance of *Franke,* the best (and only) choccocino coffee machine around. Thanks also to the EASTBIO DT Programme that economically supported my studies.

There are a few marvellous individuals I need to give a special mention and thanks to.

To Dr. Luca Tosti, my first mentor, for teaching me and believing in me even when I didn't put primers in the PCR mix.

To Dan, my partner in sass who taught me the Kaji lab ways, not least the preparation of secret Santa survival kits.

To Meryam, the best "stinky" French friend I could ever ask for. Inshallah, we'll get to share more life drawing and bao buns together.

To the PhD ladies Monica, Martha, Azzurra, Vasiliki, and all the other students who shared the ups and downs of SCRM life. To Mitzy, for all the amazing Disney songs we sang while preparing lentiviruses, all the library hours listening to the *silent* disco, and for being my food stall partner.

To my Italian family and friends, for supporting me through time and distance. To Giulia, to Emanuela and to Lorena, for being my anchors and roots.

To Luca, for sharing the journey with me, fighting sock serial killers and guiding me through dark places and sushi buffets.

To Alessio, for his trust and great patience. And daily dog gifs.

To Pedro, for being the most wonderful dog, too cute to get scolded for his mischief.

To Mum and Dad, for their big hearts and their infinite support, and for overcoming the heartache of distance by stuffing parcels with pastina and gnomes.

To Cecilia, for every message, for every update, and for her strength. For understanding. And for making me proud to be her sister.

Abstract

As the global phenomenon of population aging has become very prominent in the last few decades, diseases of the aging such as Alzheimer's and Parkinson's disease have taken the spotlight as the new challenges of modern medicine. In fact, aging is the biggest risk factor for neurodegeneration.

Neurodegenerative diseases are particularly difficult to study, as animal models often fail to recapitulate human disease phenotypes, together with the challenging aspects of monitoring cognitive functions in smaller animals.
Thanks to stem cell research, we're now able to derive large quantities of human neurons *in vitro*. This has been incredibly helpful for neuroscientific research, but stem cell-derived neurons show cellular and molecular features, such as epigenetic characteristics and metabolic activity, that resemble embryonic neurons and thus fail to recapitulate the damage and oxidative stress typical of aged neurons.

In 2010 the first report of *in vitro* fibroblast-derived neurons opened the way for an alternative source of neurons, that could easily be derived from individual patients, and most importantly maintain the cellular age of the original fibroblasts. This presents then as a better tool to study diseases that affect aged neurons *in vitro*. The conversion process is nonetheless quite inefficient and leads to a very heterogeneous mix of cell types, in which a limited number of *induced* neurons would need to be isolated from contaminating fibroblasts, myoblasts and cells of uncharacterized subtypes.

In this thesis I describe my efforts to identify the drivers of the reprogramming process that lead to a successful conversion from a fibroblast to a neuron, and the roadblocks that might be hindering it. To achieve this, I decided to take a genetic screening approach, taking advantage of the emerging CRISPR/*Cas9* technology that has made genetic knock-outs relatively easy and quick to achieve. By converting

Cas9-expessing mouse embryonic fibroblasts (MEFs) into induced neurons (iN) while expressing target gRNAs, I performed a loss-of-function screen to identify genes involved in the conversion process.

I first identified a genetic reporter that allows me to isolate the induced neurons from the remaining cell types in the heterogeneous culture. I then selected a subset of genes within the mouse genome as likely candidates to be involved in the cell identity conversion and then proceeded to perform the screen.

I identified two genes, *Stxbp1* and *Sf3a1*, as required for iN conversion and survival in the conversion context. *Stxbp1* is known to be required for neuronal survival, as its knock-out prevents neurotransmitter release and leads to neuronal death. This thus represents confirmation that iN are a reliable platform to study neuron biology. *Sf3a1* is part of the SF3A complex involved in pre-mRNA splicing and has previously been reported to be required for cell identity transitions, but, to the best of my knowledge, had yet not been associated with cell survival in the context of fibroblast to neuron conversion.

This work presents a solid platform to genetically engineer and characterize *in vitro* induced neurons and highlights their relevance to broader neuronal culture systems and general neuron biology.

Lay summary

Neurodegenerative diseases have been emerging as the next big challenge in modern medicine, as the increase in average life expectancy has led to global population aging.
These diseases are associated with defects of cognitive and motor functions as the brain and nerves decline over time.
It is crucial to understand what keeps brains and neurons healthy, what goes wrong in the cells of patients affected by disease, and how to address these defects to restore proper functions.

Neurodegenerative diseases are particularly difficult to study, as very few animal models can recapitulate the pattern of disease seen in humans, and culturing neurons in a lab is challenging. Advances in research around stem cells – the cells with the potential to grow indefinitely and generate any cell type in the body – meant we can now derive and characterise large quantities of human neurons in a lab. Unfortunately, the neurons derived from stem cells resemble the neurons of embryos and young children, and thus are not a good model to study diseases of aging, where neurons have accumulated internal damage and stress over the years.

A few years ago, scientists managed to convert human skin cells into neurons in the lab. The neurons obtained from skin cells actually maintain the “cellular age” of the patient from whom the cells were originally taken. This provides an exciting platform to study aged neurons and understand what goes wrong in cells of patients with neurodegenerative diseases. Nonetheless, the efficiency of converting skin cells into neurons is very low, and many of the cells remain stuck at an intermediate stage or die off.

In my PhD, I investigated what happens during a successful conversion of a skin cell into a neuron, and what genetic mechanisms are responsible instead when the conversion fails.

By eliminating the genetic components that hinder the conversion, and thus efficiently producing a pure population of neurons in a dish, this system could one day be used in the clinic to study patient-derived cells and potentially tailor drug treatments based on the response of the patient's own cells.

Table of contents

List of figures

List of tables

List of abbreviations

BFP	Blue fluorescent protein
BAM	Brn2, Ascl1 and Myt1l
bHLH	Basic helix-loop-helix
ChIPseq	Chromatin Immunoprecipitation followed by Next Generation Sequencing
CRISPR-Cas	Clustered regularly interspaced short palindromic sequences and associated Cas proteins
crRNA	CRISPR RNA
DSB	Double Strand Break
dsDNA	Double stranded DNA
ESC	Embryonic stem cell
FDR	False discovery rate
gDNA	Genomic DNA
GO	Gene ontology
iN	Induced neuron
iPSC	Induced pluripotent stem cell
KD	Knock-down
KO	Knock-out
MAGeCK	Model-based Analysis of Genome-wide CRISPR/Cas9
MEF	Mouse embryonic fibroblast
MOI	Multiplicity of Infection
NGS	Next generation sequencing
NHEJ	Non homologous end joining
NLS	Nuclear localisation signal
pA	Polyadenylation signal
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
RNAi	RNA interference
sgRNA	single guide RNA
TALEN	Transcription activator-like effector nuclease
TF	Transcription factor
TracrRNA	trans-activating RNA
TSS	Transcriptional start site
Wt	Wild type
ZF	Zinc finger

Chapter 1 Introduction

1.1 Reprogramming cellular identity *in vitro* to study specialised cell types

1.1.1 A historical note on cell identity and cell fate changes

The National Institutes of Health (NIH) claim that our bodies contain circa 200 different cell types. This has long since been disputed, as the technological advances in imaging and single-cell "-omics" gave us a more comprehensive understanding of the detailed composition of each cell. Much like the Human Protein Atlas Project (Uhlén *et al.*, 2005; Uhlen *et al.*, 2015), several research groups and funding agencies, including the Wellcome Trust and the Chan-Zuckerberg Initiative, have set-out to define a Human Cell Atlas and compile a map of all the cell types and subtypes that make-up the tissues in the body (Yong, 2016).

Crucially though, even though all these cells can have hugely different functions, they all share the same genetic information (DNA). This was a revolutionary discovery during the 20th century, as until then, one of the hypotheses around embryology and development was that cells progressively lose genetic information as they specialise and only retain the genes necessary for their function.

The first steps in characterising this phenomenon came from the work of developmental biologists Robert Briggs and Thomas King. Using *Rana pipiens* as a model organism and a nuclear transfer technology they developed, they transplanted nuclei from blastula, gastrula and tail-bud cells into enucleated egg cells. Blastula and gastrula cell nuclei were able to generate viable offsprings, although gastrula nuclei showed lower efficiencies, whereas tail-bud cells initiated embryonic development but induced severe abnormalities (Briggs & King, 1952).

Building on this work, John Gurdon in 1962 performed a series of experiments on *Xenopus laevis* that would eventually lead him to be awarded the Nobel Prize in 2012. Firstly, he showed that, after optimisation of the nuclear transfer technique, he could

use nuclei from tadpole gut epithelium as donor nuclei for successful development of embryos, initially up to the larva stage (Gurdon, 1962a) and then into fully developed and fertile individuals (Gurdon, 1962b). This was proof that fully differentiated cells contain all the genetic information to produce a whole organism. It was also proof that cell identity is not set in stone: in this case, nuclear identity could be reprogrammed by the egg cell environment, to a zygote state.

Only a few years before, in 1957, Conrad Waddington had described in his famous landscape model the idea that during development, as cells divide they go through a series of "decisions", as sliding down the slopes of a valley, that ultimately lead them to a precise spot, and can't go back upwards (Waddington, 1957) (Fig.1.1).
Briggs, King and Gurdon's works represent the first dispute to the verticality of this model and suggest how cell fate decisions have the potential to be reversed and move up the hypothetical valley.

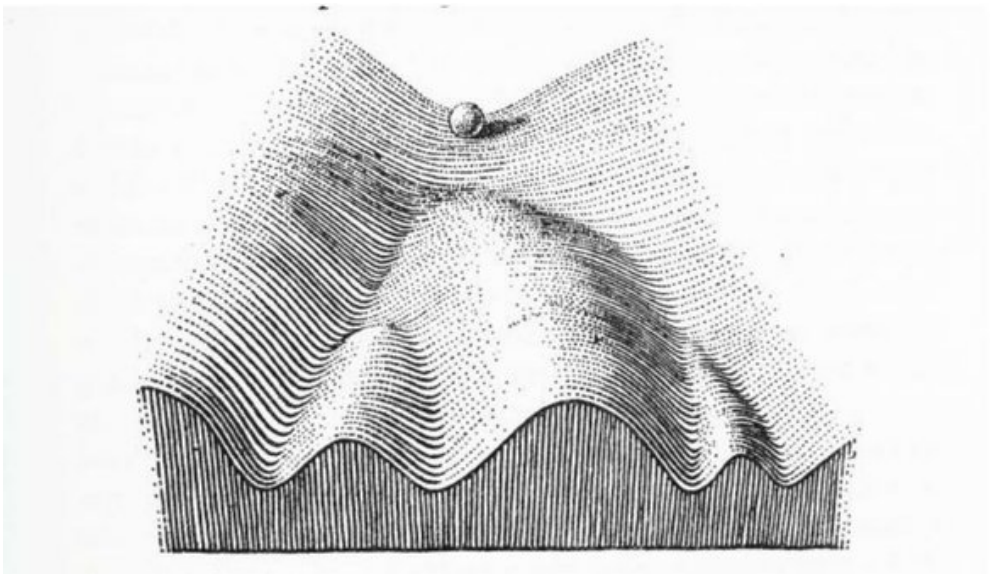


Fig. 1.1 – Waddington landscape model of development. (Taken from Waddington, 1957.) Epigenetics idea that cells restrict their potential as sliding down a valley and landing in different positions based on the "decisions" taken down the "developmental valley" forks.

Gurdon's work was reproduced in mammalian cells, first famously in sheep with the birth of Dolly, the lamb cloned from the nucleus of mammary gland cell (Wilmut *et*

al., 1997), then subsequently in mice (Wakayama *et al.*, 1998). The relatively low rates of success from these studies lead some people to speculate that successful nuclei-transfer reprogramming relied on the involuntary selection of rare stem cell populations as nuclei donors. The final confirmation that fully differentiated, somatic nuclei transfer can generate fully functional adult organisms came from a work using mouse lymphocytes as donor cells. As each clonal B-cell line in the adult organism rearranges its immunoglobulin locus in specific and traceable combination, the authors were able to show that mice generated from B-cell nuclei showed, in every tissue examined, the exact genomic re-arrangements of the nuclei of origin (Hochedlinger & Jaenisch, 2002).

1.1.2 Transcription factors regulate gene expression and can reprogram cell fate

The nuclear reprogramming studies suggested that some factors contained within the egg cell cytoplasm were able to convert differentiated nuclei to a zygote-like state. To try and understand this phenomenon and figure out whether only egg cells possess this capacity, scientists attempted to fuse cells together using polyethylene glycol (PEG) to induce membrane fusion. Blau and colleagues used this technology to fuse together human amniocytes with mouse muscle cells and observed the formation of a heterokaryon made of multinucleated, non-dividing cells. After 24h, the human amniocyte nuclei had upregulated the expression of human muscle genes, and human muscle proteins were detected in the shared cytoplasm thanks to species-specific antibodies (Blau, Chiu & Webster, 1983). This work confirmed the existence of cell-type specific factors that can reprogram exogenous nuclei.

4 years later, two studies were able to show that the overexpression of a single transcription factor can alter cell fate. In *Drosophila*, the overexpression of the *Antennapedia* gene at specific developmental stages altered the body segmentation and induced the formation of legs in place of antennae (Schneuwly, Klemenz & Gehring, 1987). Davis and colleagues were instead studying the *in vitro* conversion of fibroblasts into myoblasts by the addition of 5-azacytidine. They identified 3 genes

that were uniquely and highly expressed in myoblasts and showed that the single overexpression of one of those, *MyoD*, could force 50% of the fibroblasts to be converted into myoblasts (Davis, Weintraub & Lassar, 1987).

Several studies further showed how cell identity can be redirected by gene overexpression. In 2000, lymphoid progenitor cells were converted into myeloid progenitors by exogenous expression of IL-2 and GM-CSF (Kondo *et al.*, 2000). In 2004, B-cells were converted into macrophages by the overexpression of CEBP-α and CEBP-β (Xie *et al.*, 2004).

1.1.3 Reprogramming fibroblasts to a pluripotent stem cell state

The previous studies had shown that nuclear transfer, cell-cell fusion or gene overexpression can reprogram differentiated cells into different types. It was unclear though which factors within oocyte cytoplasm could convert a differentiated nucleus into a pluripotent-like state. Takahashi and Yamanaka addressed this by compiling a list of 24 genes associated with pluripotency, self-renewal and proliferation in ES cells and tumour cells, and overexpressed them as a pool in mouse embryonic fibroblasts, showing that they could successfully generate cells with ES-like characteristics, termed induced pluripotent stem cells (iPSCs). By sequential depletion of each of the 24 factors, they identified the combination of the 4 factors *Oct4*, *Sox2*, *Klf4* and c-*Myc* as the minimal combination for successful iPS reprogramming (Takahashi & Yamanaka, 2006). While the first generation-iPSCs were not able to make live chimera mice, in 2007 generation of iPSCs with functions equivalent to ES cells was demonstrated by the formation of chimeric organisms, contribution of IPSCs to the germ layer, and the generation of live and healthy embryos that were entirely iPSC-derived via tetraploid complementation (Okita, Ichisaka & Yamanaka, 2007; Wernig *et al.*, 2007).

The same result was achieved in human cells just one year after the first mouse iPSCs (Takahashi *et al.*, 2007). This discovery, the ability to produce pluripotent cells that can make any cell type *in vitro* and the potential applications of these cells *in vivo*,

opened up huge possibilities for clinical research, and opened up a whole new basic research field. John Gurdon and Shinya Yamanaka were awarded the Nobel Prize for Physiology and Medicine for "*the discovery that mature cells can be reprogrammed to become pluripotent*", or as Takahashi described in a review paper "*lowering gravity on Waddington's epigenetic landscape*" (Fig.1.2) (Takahashi, 2012).

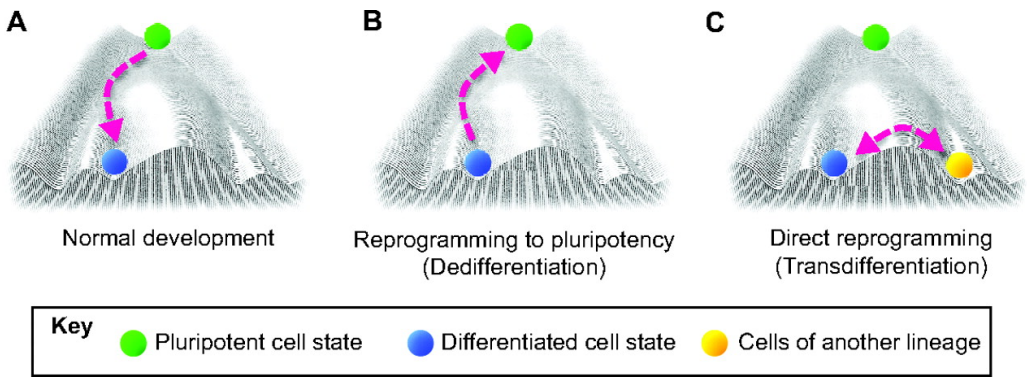


Fig. 1.2 – Takahashi's revised Waddington model for cell identity transitions.

(Taken from Takahashi, 2012). (a) The traditional Waddington model follows cell identity specialization downwards through a valley. The choices made at the forks will determine the cell fate – indicated as the position of the cell at the end of the differentiation (from green to blue). (b) In reprogramming to pluripotency, cells are brought back up the slope to the original state where cells still have the potential to go down any of the paths. (c) In conversion, like the conversion of fibroblasts to myoblasts or B-cells to macrophages (and indeed fibroblasts to neurons as discussed in the next paragraph), cells are able to jump between two crooks in the valley bed without the need to go up the slope. Overall, this shows that cell identity is a lot more fluid and changeable than previously thought.

1.2 Induced neurons: overview and implications for clinical applications

1.2.1 The first reported conversion from fibroblasts into induced neurons *in vitro*

Following the work by Takahashi and Yamanaka, in 2010 the Thomas Vierbuchen from Marius Wernig lab took a similar approach to identify factors that could convert fibroblasts into neurons *in vitro* (Vierbuchen *et al.*, 2010). A pool of 19 genes involved in neural development and epigenetic reprogramming was transduced into MEFs via

lentiviruses. To identify successful conversion and the emergence of neurons, they used MEFs from a *Tau*-eGFP knock-in mouse line that had previously been shown to selectively activate eGFP expression in neurons (Wernig *et al.*, 2002). 12 days post-transduction, the authors identified *Tau*-eGFP +ve cells with neuronal morphology that were also expressing *Tubb3*, a pan-neuronal marker. This indicated that some combination of factors among the 19 was able to reprogram fibroblasts into what they defined as "induced neurons" (iNs) (Fig.1.3a).

To narrow down the responsible genes, they first tested the transcription factors *Ascl1* and *NeuroD1* individually and showed how *Ascl1* is able, although at much lower efficiencies and generating extremely simple morphologies, to reprogram MEFs into iNs. They then proceeded to test *Ascl1* in combination with each of the 18 remaining candidates and showed how *Brn2*, *Brn4*, *Myt1l*, *Zic1* and *Olig2* increased the number of TUBB3 +ve cells (compared to *Ascl1* alone) by at least 3-fold, though failed to induce neuronal reprogramming when transduced individually. Thus, these 5 genes represented potential conversion facilitators.

After excluding *Brn4* for its similarities to *Brn2*, the authors defined a 5F pool made of *Ascl1*, *Brn2*, *Myt1l*, *Zic1* and *Olig2* and showed how this pool could generate within 12 days post-transduction cells with complex neuronal morphology expressing *Tau*-eGFP, TUBB3, MAP2, NEUN and SYNAPSIN1 (Fig.1.3b). To identify the genes strictly sufficient for the conversion, genes within the 5F pool were further narrowed down, until the combination of *Ascl1*, *Brn2* and *Myt1l* was shown to be the most efficient in generating the highest number of neurons and the most complex neuronal morphologies (Fig.1.3c). Neurons obtained from this combination, referred to as "BAM", showed the same neuronal markers as the 5F pool converted cells. They could also, upon astrocyte co-culture, form functional synapses, detected by excitatory postsynaptic currents (EPSCs) mediated by NMDA and AMPA receptors. Inhibitory postsynaptic currents were not detected in any of the neurons assessed. 53% of these neurons expressed *Tbr1*, a marker of excitatory cortical neurons.

The BAM combination was also able to efficiently generate neurons, with the same properties listed above, starting from adult tail tip fibroblasts, indicating the

robustness of this combination. Importantly, Marro and colleagues in 2011 were able to convert terminally differentiated hepatocytes into iNs using the same factors, indicating that their efficiency does not depend on MEF subpopulations or fibroblast-specific chromatin environment (Marro *et al.*, 2011).

In terms of efficiency, the BAM combination gave the best neuronal conversion efficiencies. Neurons were counted at day 12 post-transduction as TUBB3 +ve cells with a thin process at least 3 times longer than the cell soma. The average of 30 randomly-selected fields was used to estimate the total number of neurons in the dish. The ratio between the total number of neurons yielded at day 12 and the initially seeded MEFs offers a measure of the conversion efficiency. The BAM factors yielded a conversion efficiency of about 20%, with 5% variability across technical replicates.

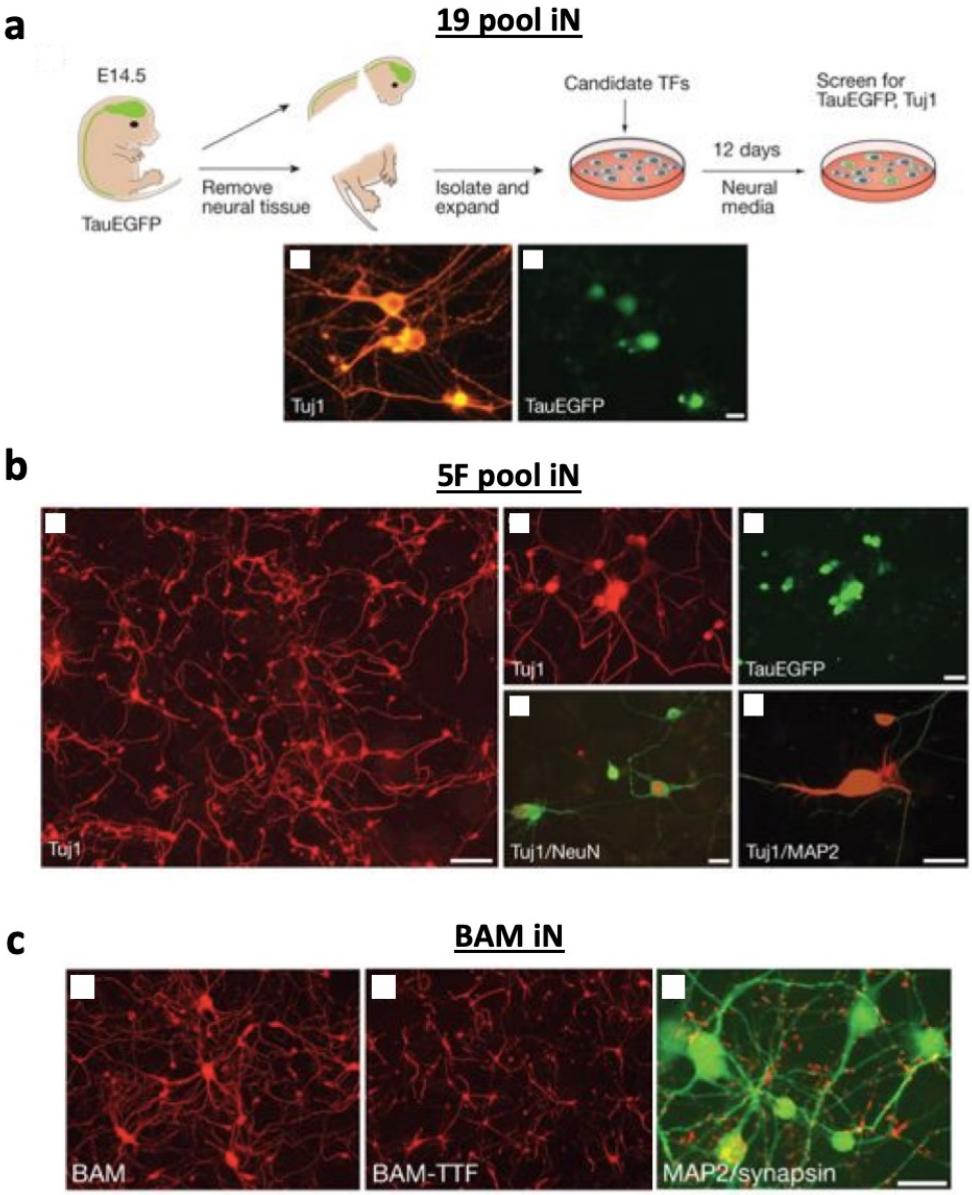


Fig. 1.3 – The first reported induced neurons. (Adapted from Vierbuchen *et al.*, 2010). **(a)** *Tau*-eGFP MEFs from non-neural tissues were transduced with a pool of 19 neural-related genes and epigenetic regulators and kept in neural media to induce the conversion of MEFs into neurons. The presence of neurons was assessed by eGFP fluorescence and *Tubb3* expression. The panels below show day 12 iNs expressing Tubb3 (left) and *Tau*-eGFP (right). **(b)** 5F derived iNs show complex morphologies and express TUBB3, *Tau*-eGFP, NeuN and MAP2. **(c)** BAM derived iNs express TUBB3 and show complex morphologies whether they're derived from MEFs (left panel) or TTFs (middle panel). They also express MAP2 (in green) and SYNAPSIN 1 (in red) (right panel).

1.2.2 Transdifferentiated iNs were generated from different species, sources and into different neuronal subtypes

One year later, a review paper by the Wernig lab provided a summary of the recent publications in the newly born induced neuron field (Fig.1.4).

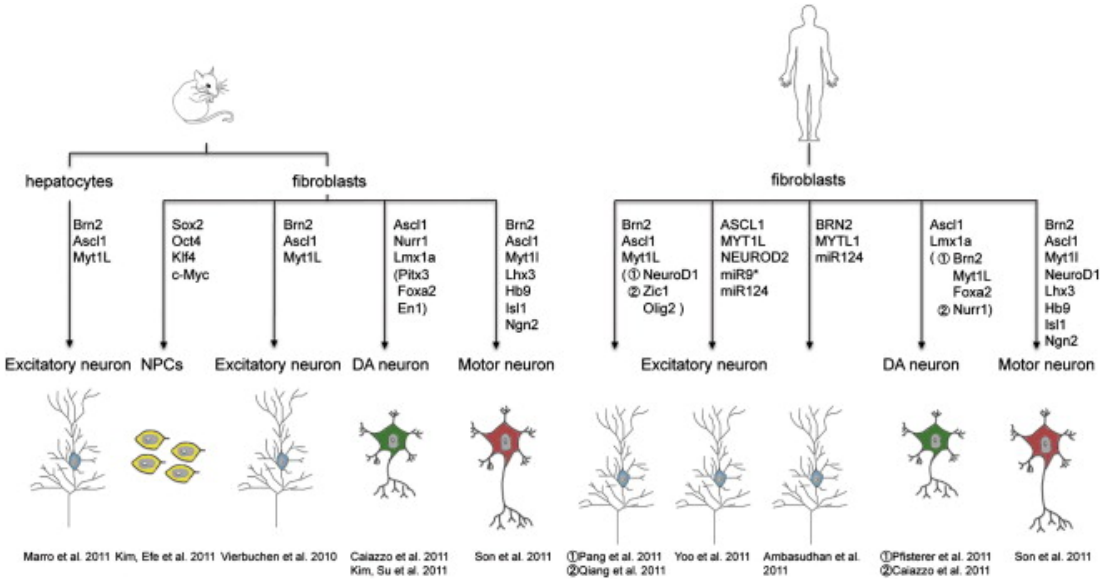


Fig. 1.4 – A few important iN publications, one year after the initial reported conversion. (Taken from Yang *et al.*, 2011). This diagram highlights a few key publications reporting the forced conversion of neurons (or neural cells like NSCs) published just one year after the initial report by Vierbuchen and colleagues. It shows how the conversion was quickly achieved for human cells too, and how different factor combinations were adapted to generate specific neuronal subtypes.

The first major point to highlight is that iNs were readily generated from human fibroblasts, but the BAM cocktail was shown to generate few, immature neurons and thus required the supplementation of extra factors. Pang and colleagues added *NeuroD1* and showed that the neurons generated could indeed produce action potentials, and upon primary cortical neuron co-culture, generated spontaneous and evoked postsynaptic currents (Pang *et al.*, 2011). Qiang and colleagues instead obtained similar results using the 5F pool cocktail of the Vierbuchen paper (Qiang *et al.*, 2011). Both studies focused on human embryonic or postnatal fibroblasts. Later

on in the same year, Pfisterer and colleagues demonstrated that human iNs can be obtained, using the BAM cocktail alone, from distal lung fibroblasts of individuals aged between 23 and 65 years old, with similar efficiencies (2-4% of the initially seeded fibroblasts) compared to the embryonic and postnatal cells (Pfisterer *et al.*, 2011b). This highlighted the potential of these neurons to be used for personalised medicine, for both patient-specific disease modelling, but also potentially cell-based therapeutics. Importantly for this regard, in 2018 human iNs were derived from adult peripheral blood cells (T-lymphocytes), with similar efficiencies to the skin biopsy-derived fibroblasts, making the derivation of patient-specific iNs in sufficient quantities even more simple and direct (Tanabe *et al.*, 2018).

The second major point is the observation that minor tweaks and additions to the reprogramming cocktail, for both mouse and human cells, allowed the derivation of specific neuronal subtypes. Dopaminergic neurons were obtained for instance by factors that specify their ventral mid-brain *in vivo* identity like *Lmx1a* and *Foxa2*, and have particular relevance for *in vitro* modelling and potential cell replacement therapies for Parkinson's disease, the second most common neurodegenerative disease nowadays (Caiazzo *et al.*, 2011; Pfisterer *et al.*, 2011a).
Similarly, induced motor neurons were obtained from both mouse and human fibroblasts by supplementing the BAM cocktail with motor neuron specific factors (Son *et al.*, 2011). They present both as potential research platform and clinical applications for motor neuron diseases such as amyotrophic lateral sclerosis (ALS) or spinal muscular atrophy (SMA).

Induced dopaminergic neurons and motor neurons are just two examples of neuronal subtypes that have been since derived and characterised.
Fig. 1.5 shows an overview of the different neuronal subtypes described in induced neurons *in vitro* and their master regulatory network. As reviewed by Masserdotti and colleagues (Masserdotti, Gascón & Götz, 2016), the search subtype specification has not only focused on supplementing the original BAM combination with regional

or neurotransmitter regulators, but has looked for alternative reprogramming factors altogether. *Ascl1* and *Neurog2* (*Ngn2)* are responsible for the differentiation of GABAergic and glutamatergic neurons from the ventral and dorsal neural progenitors of the telencephalon respectively, and have been shown to activate distinct neuronal gene networks (Pang *et al.*, 2011; Wapinski *et al.*, 2013; Zhang *et al.*, 2013; Smith *et al.*, 2016; reviewed in Colasante *et al.*, 2019). *Ascl1* represents in fact the main driver of the conversion within the BAM combination, as it acts as a pioneer TF and can independently generate iNs (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013; Treutlein *et al.*, 2016). *Ngn2* has also been extensively used as a main driver to derive iN from mouse and human pluripotent stem cells, although it's unable to reprogram fibroblasts by itself (Chanda *et al.*, 2014). With the help of small molecules though, it has been shown to act as a pioneer factor and is able to convert fetal and adult human fibroblasts into induced neurons (Smith *et al.*, 2016).

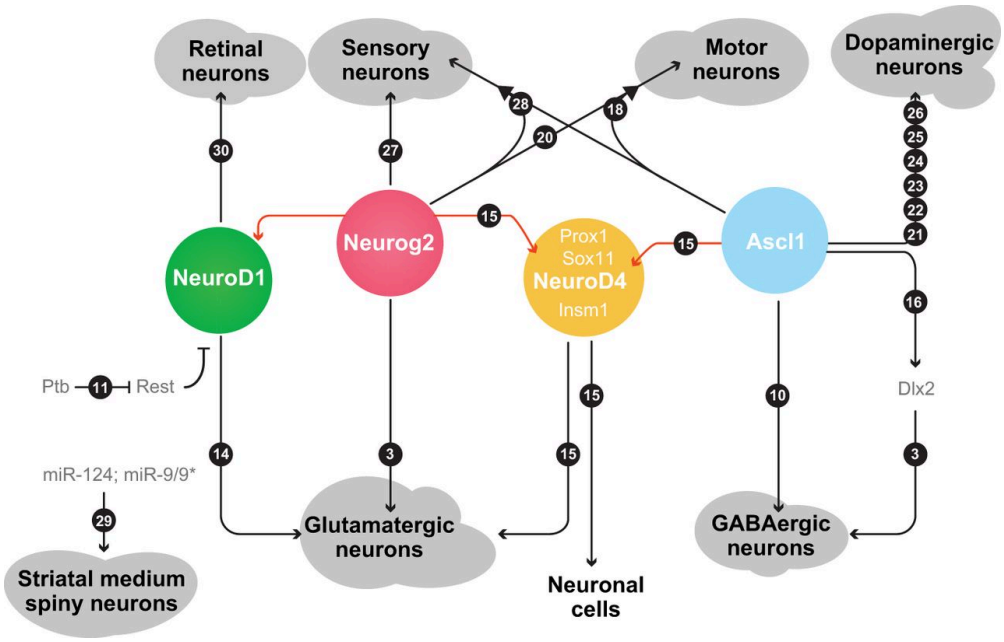


Fig. 1.5 – Induced neuron subtype specification. (Taken from Masserdotti, Gascón and Götz, 2016). The coloured nodes indicate the 4 major TF used as master regulators of induced neurons *in vitro*. Combination of other factors and/or miRNAs has been shown to direct neuronal subtype specification – indicated in the grey bobbles. The numbers reported in black circles refer to the reference numbers in the review paper that report the specified neuronal subtype generation.

1.2.3 Neuron reprogramming *in vivo* – implications for cell replacement therapies

For the purpose of cell replacement therapies, direct conversion of neurons represents a particularly interesting system as it doesn't pass nor require a proliferative stage (Fishman *et al.*, 2015; Gascón *et al.*, 2016) and it's thus unlikely to pose a tumorigenic threat if cells are transplanted or converted *in vivo.* In 2013 Torper and colleagues showed that human fibroblasts could be transduced with lentiviruses containing the tetO-BAM combination and transplanted within rat brains to achieve engraftment. By supplying doxycycline in the animals' drinking water, they activated BAM expression and achieved conversion of the transplanted human fibroblasts into iNs with similar efficiency to the *in vitro* conversion (Torper *et al.*, 2013). They also showed how dopaminergic iNs can be induced *in vivo*, and how the same iNs can be induced by transplanting human astrocytes. Following on this work, they showed they could convert endogenous mouse brain astrocytes into induced neurons *in vivo* by delivering the BAM factors lentivirally. This publication highlighted the potential of *in vivo* reprogramming as a strategy for brain repair, and set-out the question of finding the right combination and balance of factors that would induce the generation of subtype-specific neurons able to integrate within the neural network at the location of reprogramming.

A year later, Guo and colleagues reprogrammed glial cells into functional neurons *in vivo*, using retroviruses to deliver a *NeuroD1* overexpression cassette into the mouse cerebral cortex (Guo *et al.*, 2014). They demonstrated how *NeuroD1* could reprogram astrocytes into glutamatergic neurons, and NG2 glial cells into glutamatergic and GABAergic neurons, proving evidence on how different targeting strategies (using target cell specific promoters to induce TF expression) could generate in neuronal subtypes *in vivo* and could be exploited to obtain a subtype-specific pure neuronal population.

Direct generation of iN *in vivo* thus presents as a competing strategy for cellular replacement therapies for diseases of the CNS – a parallel approach to exogenous cell transplants that involve the engraftment of neural progenitors and their *in vivo* differentiation, and to endogenous brain repair that targets instead resident neural

progenitor cells to replace lost cells, or brain inflammatory cells to repair damage. All these strategies could play crucial roles in treating a whole range of neurodegenerative diseases that result in neuron loss like ALS, PD, Alzheimer's, or even injuries. After the proof of concept studies have been released, research has focused and is still focusing on optimising the target cells to convert, the neuronal subtypes to induce and integrate, the transcription factor delivery and how to make it safe (non-integrating, less invasive, or possibly substitute it with small molecules) (Barker, Götz & Parmar, 2018).

1.3 Induced neurons as a disease modelling platform

1.3.1 Modelling neurological diseases *in vitro* – mimicking complexity and cellular age

Human neurological diseases are extremely difficult to study, because of the intrinsic complexity of the human CNS, that is both hard to replicate *in vitro* and to model in animals (phenotypes like social behaviours or pain can be quite tricky to assess in mice and rats for instance).

The relatively recent development of cerebral organoids (Lancaster *et al.*, 2013), building on ESC to neuron differentiation protocols, has just begun to address one of the major hurdles of studying neurons *in vitro*, i.e. the difficulties of relating the synaptic properties and networks of isolated neurons or specific co-culture systems to the huge and coordinated networks of multiple cell types *in vivo* and in different brain regions. Organoids have been useful in modelling brain development and natural progenitor differentiation, migration and assembly into cortical layers of all neural cell types, as they've relied on pluripotent stem cells as a starting cell type. They've been also useful in reproducing disease phenotypes – by using patient-derived iPSCs, the authors showed how the cerebral organoids mimic the pathogenesis of microcephaly much better than previous problematic mouse models (Lancaster *et al.*, 2013; reviewed in Lancaster and Knoblich, 2014).

These models rely on the differentiation of pluripotent stem cells into neural progenitors and then neural cells. This necessary step introduces a big limitation in the study of neurological diseases related to aging, because the pluripotent state of ES cells and iPS cells associates them with an embryonic cellular "age". As reviewed by Studer and colleagues, upon iPSC reprogramming, the cells reset many characteristics of aged cells, including notably epigenetic changes (DNA methylation and heterochromatin state), DNA damage marks like γ-H2AX, protein aggregation state, mitochondrial DNA mutations and ROS, nuclear lamina composition and integrity, telomere length and markers of senescence (Studer, Vera & Cornacchia, 2015).

An example of their limitations to generate age-appropriate neurons for disease modelling was described in 2013: *in vitro* differentiated dopaminergic neurons derived from iPSCs from Parkinson's patients failed to show significant neurodegenerative phenotypes like cleaved CASPASE-3 in condensed nuclei or shortening of dendrite length. These phenotypes were instead induced by forced expression of progerin, the truncated form of LAMIN A that is associated with premature aging and is responsible for Hutchinson–Gilford progeria syndrome (Miller *et al.*, 2013). Similar results, again in dopaminergic neurons *in vitro*, were obtained by artificially aging hiPSCs via telomere shortening (Vera, Bosco & Studer, 2016). This suggested that induced aging might be a good tool to reveal disease-associated phenotypes in hiPSC-derived neurons, but introduced an extra, non-physiological step that might influence the relevance of *in vitro* aged neurons to *in vivo* neurons of old patients. Considering the global population is aging and neurodegenerative diseases are quickly becoming a huge burden for patients, families and healthcare providers around the world, they need to be urgently understood and addressed. It's also worth noting how human life-span is significantly different from other primates, let alone traditional animal models like rodents, and how the complex but common symptoms associated with dementia, like memory loss, confusion, loss of perception of time or space can be extremely challenging to define and monitor in animals (reviewed in Mertens *et al.*, 2018).

Induced neurons instead present as an ideal platform to model neurological aging diseases, as they have been shown to maintain the cellular age of the fibroblasts from which they are derived (Mertens *et al.*, 2015). Human fibroblast samples from patients aged 0-89 years were reprogrammed in parallel to iPSCs and iNs, confirming that iPSCs, regardless of the donor age, reset the aging hallmarks. The iN conversion instead showed how donor age does not impact conversion efficiency, but is revealed upon transcriptome analysis: iNs derived from >40 years old donors showed differentially expressed genes associated with neuron specific aging phenotypes like loss of synaptic plasticity and Ca^{2+} homeostasis. Interestingly, old iNs showed functional defects associated with aging, like leaky nuclear membrane and impaired nucleoplasm compartmentalisation.

The same effects were seen in induced motor neurons derived either directly from human fibroblasts or induced from fibroblast-derived iPSCs: direct reprogramming maintained aging features such as increase numbers of DNA-damage associated γH2AX foci, increased SA-β-Gal activity and increased mitochondrial ROS (Tang *et al.*, 2017). This was further proof that the maintenance of cellular age was not specific for a subset of reprogramming factors or neuronal subtype.

This aged phenotype was confirmed in mouse cells as well, as iNs derived from 1 year old mice tail tip fibroblasts (TTFs) showed epigenetic marks, DNA damage and stress response gene expression of aged TTFs, compared to iNs derived from differentiated iPSCs (Yang *et al.*, 2015).

A more thorough investigation made use of the “epigenetic clock”, an algorithm to predict cellular age based on DNA methylation focused on a subset of 353 key CpG sites that were determined to be good predictors of tissue donor age (Horvath, 2013). This algorithm was shown to associate iPSC epigenetic age with 0, consistent with the age-resetting already discussed. Striatal medium spiny iNs were derived from fibroblast samples aged from 3 days to 96 years old, and their epigenetic age was assessed with the described algorithm. This showed a great correlation (0.91) with the epigenetic age of the cells of origin, as well as a good correlation (0.75) with the

patient's actual age, indicating that, at least on a DNA methylation level, iNs show minimal to no age resetting (Huh *et al.*, 2016).

1.3.2 Induced neurons recapitulate disease pathology *in vitro*

The suitability of iNs to model diseases has been demonstrated soon after their establishment as a platform: MEF-derived iNs from mice carrying an autism-associated mutation in the *Neuroligin-3* gene showed impaired synaptic transmissions through AMPA glutamate receptors but not NMDA glutamate receptors nor GABAA receptors, consistent with the observations of the *in vivo* neurons of the adult animals (Chanda *et al.*, 2013). This report also strengthened the concept that iN, even as a heterogeneous mix of neuronal subtypes derived from the original BAM cocktail, are functionally similar enough to *in vivo* neurons to allow the identification and characterisation of disease-specific changes at the synaptic level.
Subtype-specific neurons have also been shown to respond to disease stimuli: induced motor neurons derived from MEFs from a mouse model of ALS showed reduced survival, and wild-type induced neurons co-cultured with glial cells from the ALS models also showed reduced survival, indicating that the induced motor neurons are suitable to study both cell-autonomous and non-autonomous disease phenotypes (Son *et al.*, 2011).

1.3.3 iNs model aging diseases *in vitro*

A few publications have actually attempted to use the iN conversion to study aging associated diseases with age-appropriate neurons (reviewed in Mertens *et al.*, 2018). Of note, Jovičić and colleagues have characterised the *C9orf72* mutation associated with ALS and shown how patient derived iNs, compared to iNs derived from fibroblasts from healthy individuals, show defects in nucleocytoplasmic transport, consistent with the reported toxicity associated with the mutant forms of *C9orf72* (Jovičić *et al.*, 2015).

Another group looking at ALS focusing specifically on induced motor neurons highlighted defects of patient-derived iMNs in nuclear localisation of the mutated FUS proteins and in forming neuromuscular junctions (Liu, Zang & Zhang, 2016).

A study focusing instead on Huntington's disease looked at differences in induced striatal medium spiny neurons (MSNs) derived directly from patient's fibroblasts, or induced from fibroblast-derived iPSCs (Victor *et al.*, 2018). They addressed this because previous iPSC-derived neurons failed to show protein aggregates and neuronal death typical of the disease unless artificially exposed to stressors like proteasome inhibitors. Directly reprogrammed MSNs showed the formation of huntingtin's aggregates, DNA damage, declined mitochondrial function and spontaneous cell death – features that iPSC-derived MSNs from the same patients' original fibroblasts failed to reproduce. They investigated further this disease model and found that, consistent with normal disease progression in patients, iMSNs from pre-symptomatic individuals carrying the HD mutation showed much milder disease phenotypes. Moreover, they were also able to show that induced cortical neurons derived from the same patients show much milder degeneration compared to iMSNs, consistent again with disease pathology *in vivo*. This work overall highlights how direct neuronal reprogramming can be a powerful and precise platform to model neurodegenerative diseases with age-appropriate, subtype appropriate cells.

1.3.4 Bringing iNs to the clinic: overcoming limitations of timing and cell numbers

As discussed by Mertens and colleagues, there are a few major drawbacks that limit the suitability of this system for personalised medicine and disease modelling applications. One consideration is timing: though it bypasses the iPSC reprogramming stage, it can still take 2-4 weeks for iNs to be able to generate action potentials, and 5-6 weeks more with astrocyte co-culture to show more mature electrophysiological features like spontaneous firing and postsynaptic currents. Some efforts in accelerating this maturation process have resulted in iNs showing spontaneous firing

within 25 days, though it remains to be examined if this is limited to a subset of the neurons in culture (Ruetz *et al.*, 2017).

Another major limitation is cell numbers: whereas iPSC-derived neurons are derived from an infinitely expandable source, iNs are derived most commonly from adult fibroblasts, cells with limited expansion capacity as they can only undergo a certain amount of cell divisions before they become senescent and stop dividing. The limited expansion capacity of adult fibroblasts also makes them unsuitable for stable genome engineering for the purpose of modelling genetic disease phenotypes, as most stable manipulations require an expandable and clonal-derived cell line.

Different groups have focused their efforts towards increasing neuronal yields and conversion efficiencies. Pereira and colleagues showed how delaying TF activation, i.e. allowing transduced cells to proliferate for up to 5 days before inducing TF expression, could artificially boost iN yields without the need to harvest more cells from biopsy samples (Pereira *et al.*, 2014). They also showed how a cocktail of small molecules that inhibit SMAD signalling and activate Wnt signalling could significantly increase conversion efficiency and neuronal purity, while the combination of TF activation delay and small molecule treatment resulted in final iN yields greater than the originally seeded cells and a neuronal purity of over 20%. Further characterisation of the system, removal of genetic barriers and optimisation of culture conditions have improved available protocols, resulting in neuronal purities up to 50% and the estimated potential to derive 10 billion neurons from a single biopsy (Drouin-Ouellet *et al.*, 2017).

The derivation of iNs from blood derived T-lymphocytes (Tanabe *et al.*, 2018) with comparable efficiencies to fibroblast derived human iNs (circa 6%) allowed the generation of circa 50’000 neurons from 1 ml of blood without need for cell expansion. The ease of access to patients’ blood samples compared to invasive biopsies strengthens the suitability of the iN platform for clinical investigations.

While these studies have shown great promise, a greater understanding of the molecular mechanisms involved in the conversion process would facilitate the

identification of genetic barriers and tackle issues of heterogenicity between cells, as well as donor intrinsic variabilities, and would bring us closer to using iNs as a routine patient resource.

1.4 Understanding and improving the iN system

1.4.1 ChIPseq identifies *Ascl1* as the pioneer TF driving the conversion

It is difficult to obtain a genetic and molecular understanding of a process that generates diverse neuronal subtypes, starting from different cell types and using different reprogramming systems. For simplicity and relevance to this thesis, I will discuss the efforts that have led to the characterisation of the conversion of fibroblasts into pan-neurons using the BAM combination and *Ascl1* as the pioneer TF. One of the earliest attempts at understanding the mechanisms underlying the conversion came from the Wernig lab itself. In 2013, they used ChIPseq and RNAseq to study different cell populations at day 0, day 2, day 13 and day 22 of the conversion (Wapinski *et al.*, 2013). Gene expression at 48h in *Ascl1* and BAM-transduced MEFs revealed how *Ascl1* is the main driver of the conversion and acts as a transcriptional activator, by upregulating, as early as day 2, genes related to neuronal activity. ChIPseq analysis revealed that *Ascl1* has the most binding sites within 48 BAM-transduced MEFs, and that it doesn't require the other two factors to interact with the genome. It is able to bind closed chromatin regions and thus presents itself as a pioneer TF. Interestingly, the loci targeted in MEFs are the same physiological targets that *Ascl1* binds to in neural stem cells. In contrast, *Brn2* in MEFs targets different sites compared to its endogenous loci in neural stem cells, and seems to be recruited by *Ascl1* to help with chromatin remodelling and gene activation.

1.4.2 ATAC-seq shows major chromatin accessibility changes at 48h and 5 days of the conversion

The fundamental role of *Ascl1* in the conversion was confirmed using ATAC-seq to look at chromatin accessibility dynamics during the early stages of MEF to iN reprogramming. Upon *Ascl1* transduction alone, chromatin remodelling happens extremely quickly, with significant changes at *Ascl1*-bound loci detectable 12h after TF induction. At 48h there is a clear picture: *Ascl1* bound sites are now in an open chromatin state, and associated genes are upregulated, consistent with *Ascl1* role as a transcriptional activator. Interestingly, chromatin opening regions are associated with neuron but also muscle GO terms. This is consistent with *Ascl1* and *MyoD* both binding E-boxes motifs and indicates how *Ascl1* might be promiscuously promoting a myoblast-favourable chromatin environment.

At day 5, a major chromatin conformation switch is detected by ATAC-seq, and regulatory elements now accessible are associated with genes for neuronal processes and synaptic maturation. This precedes morphological and functional changes in the neurons, that are detected significantly by day 12 onwards (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013). Sites that have instead become inaccessible relate to broader biological functions and extracellular matrix processes, closer to the profile of the original fibroblastic identity. Interestingly, day 5 cells that fail to become neurons (as assessed by the *Tau*-eGFP reporter) show an accessibility profile in between initial MEFs and day 5 iNs, as they fail to open regions associated with genes involved in synaptic maturation but instead maintain a fibroblast-like chromatin environment. As highlighted in Fig.1.6, the overall picture is that the iN reprogramming is happening relatively quickly in two crucial waves. Firstly, *Ascl1* primes its target sites for neuronal induction within the first 48h. It then recruits the other factors, and by day 5 gene regulatory elements associated with neuronal maturation become accessible. This leads to morphological and functional stabilisation of *Tau*-eGFP +ve cells, that in fact acquire progressive neuronal features from this timepoint onwards.

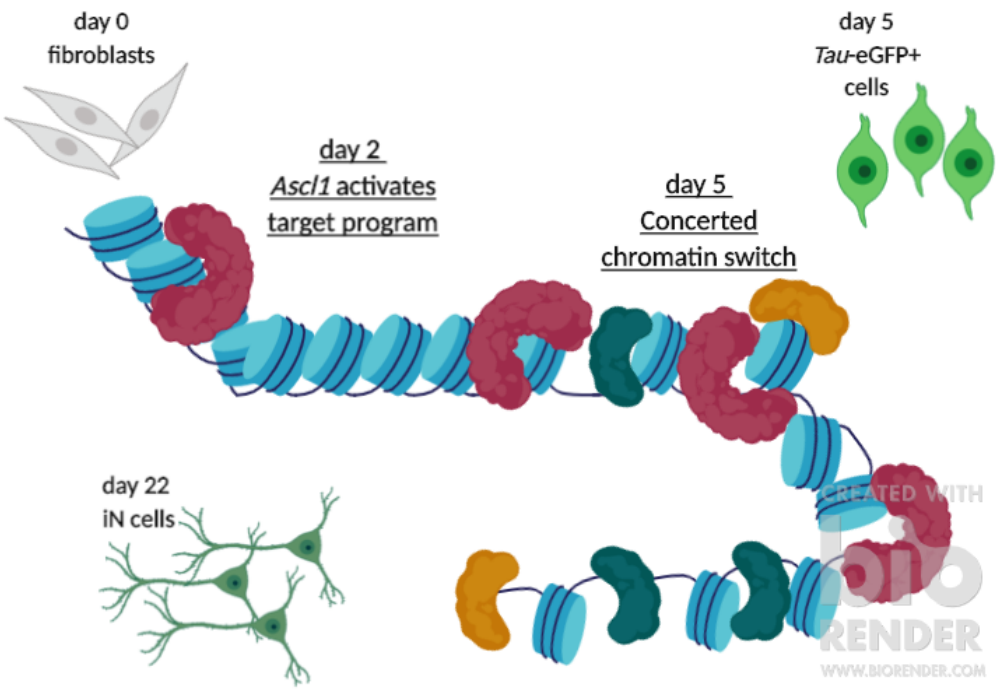


Fig. 1.6 – the pioneer TF *Ascl1* drives gene activation in the MEF to iN conversion in 2 major waves. (Adapted from Wapinski *et al.*, 2017)**.** *Ascl1* recognises trivalent chromatin state on its physiological sites in MEFs leading to conformational changes and gene activation (Wapinski *et al.*, 2013). The first wave of chromatin accessibility remodelling happens within 48h with the activation of the *Ascl1* target program. The other factors are then recruited to enhance the transcription of genes required for the stabilisation of neuronal identity, which occurs at day 5 and precedes morphological and functional changes in the iNs, which eventually reach full maturity by day 22.

1.4.3 Single cell gene expression identifies cell subpopulations and alternate cell fates

With the advances and optimisations of single cell RNA sequencing, a more thorough investigation of the iN conversion process allowed the identification of cell subpopulations and their dynamics (Treutlein *et al.*, 2016). *Ascl1*-only induction was sufficient to downregulate, by day 2, genes involved in cell cycle and mitosis, consistent with the notion that cells that become neurons become post-mitotic within 48h and do not progress through a proliferative stage (Vierbuchen *et al.*,

2010). Cells respond quite uniformly transcriptomically to *Ascl1* induction, indicating that the heterogenicity observed at later stages must be due to subsequent events. Upregulated genes are involved in neuronal programming, including cytoskeletal reorganisation, synaptic transmission and neurite projection development.

Investigation of cell transcriptomes at day 5 revealed how high levels of *Ascl1* expression correlate with high induction of *Tau*-eGFP reporter, and that cells that fail to reprogram underwent *Ascl1* silencing. There is also an upregulation, in *Tau*-eGFP-ve cells, of myoblast identity genes, consistent with *Ascl1* promiscuous binding described before.

Analysis of BAM-reprogrammed cells at day 22 revealed 3 distinct major subpopulations: iNs, as defined by *Tau*-eGFP expression and *Syp* expression, fibroblasts (marked by *Eln* expression) and myoblasts (defined by *Tnnc2* expression). Very few of the cells reprogrammed with *Ascl1* alone instead manage to mature to iNs, and show a predominant myogenic transcriptional profile. Consistently, *Brn2* and *Myt1l* expression was low in BAM-reprogrammed cells with myoblast expression at day 22. This confirmed the predicted roles of *Brn2* and *Myt1l* in directing and stabilising neuronal identity.

Interestingly, they also identified intermediate, transient cell populations that upregulate neural progenitor associated genes like *Nestin* and *Sox9* early on in the conversion, and then disappear as iNs mature. These cells never activate canonical neural stem cell markers like Sox2 or Pax6, but nonetheless, this shows the existence of transient transcriptional stages that are quite distinct from both starting MEFs and target iNs.

A further study looking at identifying *Myt1l* genomic binding sites in iNs revealed how this previously relatively uncharacterised factor acts as a transcriptional repressor and binds, in neurons and fibroblasts, to non-neuronal, open chromatin loci, fundamentally preventing the activation of any non-neuronal program, for instance by silencing the Notch signalling pathway (Mall *et al.*, 2017).

The ATAC-seq and scRNA-seq datasets were used to build a TF interaction network model that governs iN generation, placing *Ascl1* at the top of the hierarchy (Wapinski *et al.*, 2017). TFs are classified in tiers based on their expression at day 5, 13 and 22, and the way they are interconnected is based on their predicted binding sites on promoters of the other TFs – overall predicting which TFs induce and regulate which. This model was used to select 6 TFs to in MEF to iN reprogramming based on their expression levels and connectivity to *Ascl1*. Only two of those, *Sox8* and *Dlx3* were able to induce iNs when in presence of *Myt1l*, indicating that gene networks and pathway models generated from sequencing datasets won't necessarily predict gene function.

Overall, these studies provided an overview of the gene regulatory networks that orchestrate the cell identity change from a fibroblast into an induced neuron. *Ascl1* is the master regulator of the conversion, while *Brn2* and *Myt1l* stabilise neuronal fate and limit alternate routes respectively. A few effector and target genes have been identified by looking at the networks described, but their role hasn't always been functionally validated. Moreover, these types of approaches have a harder time identifying potential genes that instead prevent or slow down cell identity changes. In the next section, I'll describe how a genetic loss-of-function screening approach could address this challenge.

1.5 Genetics screens allow us to understand biological processes and the relationships between genes

1.5.1 A brief history of genetic screens

In 1993, Prof. William Sullivan, trying to explain to his students the difference between genetics and biochemistry, wrote an interesting and funny essay on the rationale behind mutational genetics. This famous piece presents two colleagues, Bill the biochemist and Doug the geneticist, observing a car factory and wondering how

all the people and materials going inside the factory every morning eventually results in automobiles driving out by the end of the day.

Compared to Bill's strenuous work of carefully taking cars apart and studying what they were made of, Doug's approach was particularly straightforward and required relatively little effort: every morning he'd tie-up the hands of one of the workers going in, and, in the evening, he would observe the cars coming out of the factory and check what they were missing (Sullivan, 1993).

Translating this into biology's terms, Doug's work is the equivalent of performing genetic knock-outs to understand the role of each gene in one specific process by impairing their function. When Doug started tying up the hands of multiple workers at once, he was, in a way, performing a genetic screen.

The first genetic screens in eukaryotes made use of yeast biology, using the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* as model organisms to characterise basic principles of eukaryotic cell biology. Thanks to their life-cycle, surviving as both haploids and diploids, they are particularly suitable to discover the phenotype of recessive mutations in their haploid form, and maintain them to allow further characterization in their diploid form (this is a particularly interesting aspect when considering performing genetic knock-out screens in mammalian, diploid, cells). Screens originally relied on chemically induced mutagenesis, introducing DNA aberration via chemical agents, followed by phenotypic screening and identification of the causal genetic mutation.

Genetic screens using chemically-mutagenized yeast strains allowed the discovery of cell division cycle genes and eventually led Leland H. Hartwell, Tim Hunt and Sir Paul Nurse to being awarded the Nobel Prize in 2001 (Hartwell, Culotti & Reidt, 1970; Nurse, 1975). Chemical mutagenesis has been extensively used for screening purposes in multicellular eukaryotes as well, and has facilitated, to mention a few important milestones, the characterisation of sexual development in *C. elegans* (Ferguson & Horvitz, 1985), of segmentation and development in *Drosophila*

(Nüsslein-Volhard & Wieschaus, 1980), and of embryogenesis in zebrafish (Driever *et al.*, 1996; Haffter *et al.*, 1996).

Although these types of classical genetic screens are useful in identifying gain-of-function mutations (that would not instead be detectable by a loss-of-function, "Doug-like" knock-out approach) and are not burdened by sequence biases, one their major hurdles is the extensive work required to map and identify the genetic mutations causing the phenotypic changes.

To overcome this, random mutagenesis technologies were developed that rely on disrupting genetic sequences by the insertion of DNA itself, using transposons or retroviruses as vectors. These approaches have been developed and optimised in mammalian cells, with a focus on mouse ES cells, with the objective of using mutagenesis screens to characterise the function of all the coding genes in the mouse genome (Hicks *et al.*, 1997). Referred to as tagged sequence mutagenesis, or gene traps, they work by delivering within cells a promoterless reporter and/or selectable marker, preceded by a splice acceptor sequence and followed by a polyA signal. This sequence gets randomly integrated within the genome and, when it gets inserted within a gene intron, leads to a fusion transcript expressed by the endogenous promoter of the gene targeted. This results in a truncated and generally non-functional version of the protein, and the expression of the reporter. The advantage is that the targeted gene can be easily identified by sequencing and mapping the loci flanking the inserted tag (reviewed in Friedel & Soriano, 2010). Nonetheless, to observe a phenotype of recessive genes, mouse lines from the gene trapped ESC clone need to be generated and crossed to obtain homozygous lines and animals.

This strategy has been used extensively, and several gene trap ES cell line banks and databases have been established (Nord *et al.*, 2006; Roma *et al.*, 2007; Taniwaki *et al.*, 2005).

As mentioned above though, revealing a recessive phenotype in mammalian cells can be particularly tricky due to their diploid nature. Although haploid ES cells have been

developed and shown to be suitable for genetic screens to allow the easy identification of recessive phenotypes, the derivation of differentiated cells often requires diploidy (Elling *et al.*, 2011; Leeb *et al.*, 2014), and thus these haploid cells have been unsuitable to derive differentiated haploid lines or primary cells, like MEFs, to perform large-scale screenings on.

1.5.2 Modern tools for gene knock-out and genome engineering

In recent years, high-throughput genetic screens have been employing the biology of RNA interference (RNAi) to achieve gene expression knock-down. RNAi was first identified as a molecular mechanism in *C. elegans* (Fire *et al.*, 1998). It is a process that involves double-stranded small RNAs targeting mRNAs based on homology and tagging them for degradation, thus effectively inducing expression knock-down, though the efficiency of knock-down itself will depend heavily on each RNA species and sequence biases. Compared to classical screens, an RNAi-based screening approach has the advantages of having a quick and easy genetic readout, as it's sufficient to sequence the mRNAs in the treatment group and identify the ones missing as responsible for the phenotype being analysed (reviewed in Boutros and Ahringer, 2008). It also crucially overcomes the need to mutagenize or trap both alleles of diploid cells to reveal recessive phenotypes, as it targets directly gene expression. The ease of use of these types of screens has facilitated unprecedented feats like *in vivo* screening or screening multiple species in parallel. Nonetheless, using this strategy for genetic screens presents one major hurdle, as it has a relatively high rate of off-target effects that increase noise when analysing datasets and increases the complexity of identifying the mRNA responsible for the phenotype. Moreover, the efficiency of knock-down can be extremely variable among small RNAs targeting the same mRNA, which makes it harder to use them as internal controls for off-target effects. Small hairpin RNA (ShRNA) expression can also interfere with miRNA expression and function as they rely and would compete for the same

processing machinery (reviewed in Mohr, Bakal and Perrimon, 2010; Mohr *et al.*, 2014).

The discovery and adaptation of zinc fingers, TALENs and CRISPR/*Cas9* systems made genetic targeted knock-out quicker, easier and cheaper than the traditional methods relying on homologous recombination, especially for mammalian models and cells. This can bypass some of the issues that knock-down poses, but ultimately answers a different biological question, as RNAi will only reveal phenotypes that are affected significantly by gene expression levels rather than the complete absence of expression resulting from knock-out. Moreover, the temporal dynamics of RNAi are also significantly different from knock-out, as the mRNA downregulation via RNAi does not impact the endogenous and untargeted mRNA transcription.

Genome engineering mediated though Zinc fingers (ZF)-nucleases (Urnov *et al.*, 2005) and TALEN (Transcription activator-like effectors nucleases) (Christian *et al.*, 2010) works on a shared principle: bringing *Fok1* endonuclease to a desired target locus by fusing the nuclease with ZF or TALE DNA binding domains with designed sequence specificity. If cells are provided with a DNA template flanked by homology arms, DBS induced by *Fok1* can be repaired by homologous recombination and can be exploited to engineer or knock-in DNA sequences within a target locus. Alternatively, cells will repair the DBS via Non Homologous End Joining (NHEJ) that will lead to a high rate of indels resulting in missense mutations by introducing premature stop codons in coding sequences of genes. (Bibikova *et al.*, 2002). This way, Zinc Finger nucleases (ZFNs) and TALENs can be used to induce gene knock-out in a way that is much quicker and easier than before (reviewed in Hsu, Lander and Zhang, 2014). ZFs and TALE modules recognise to 3bp and 1bp sequences respectively, and thus they require a multi-modular assembly in order to target specific loci effectively. Both require extensive protein engineering, which can be labour-intensive and expensive. This is one of the major reasons CRISPR/*Cas9* took over so quickly on ZFs and TALENs – as its DNA recognition relies on RNA-DNA base pairing.

1.5.3 CRISPR/*Cas9* discovery

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. They were first identified in the *iap* gene in *E.coli*, consisting of a series of 29bp repeats interspaced by five non-repetitive sequences of 32bp (Ishino *et al.*, 1987). In 2000, they were classified as a family of repeat elements widely present in roughly half of bacterial species and 90% of archaea (Mojica *et al.*, 2000), and eventually labelled CRISPR in 2002 (Jansen *et al.*, 2002). The same work highlighted the presence of well conserved genes, often adjacent to the repeat elements, and defined them as CRISPR-associated, *cas*. This gene family was initially characterised and subdivided in 3 major different systems that rely on different DNA binding and processing strategies (reviewed in Makarova *et al.*, 2011). More up-to-date classifications subdivide the CRISPR-Cas systems in two classes that contain overall 6 system types. The major differences between the classes is that class 2 systems rely on a single multidomain effector protein to induce DNA cleaveage – *Cas9* is the best characterised effector protein within class 2 (reviewed in Mohanraju *et al.*, 2016).

The evolutionary role of these sequences and genes started to become clear in 2005, when it was shown that these repeats had extrachromosomal origin (Bolotin *et al.*, 2005) and were derived from viruses (Mojica *et al.*, 2005). Mojica and colleagues also showed how viruses can't infect archaea that contain, within the repeat spacers, the virus genome. In 2007, working with *Streptococcus thermophilus* for the production of dairy products, Barrangou and colleagues performed a series of experiments to show that bacterial strains infected with bacteriophages became rapidly immune to the phage strains they were exposed to and incorporated part of the phage genome in the CRISPR1 locus by adding extra spacer sequences (Barrangou *et al.*, 2007). This represented the first direct evidence that CRISPR loci are involved in bacterial immunity. Since then, extensive characterisation of this mechanism has brought us a detailed understanding on how bacteria and archaea use CRISPR-Cas complexes as a primitive immune system (Jackson *et al.*, 2017). An overview of the major mechanisms is outlined in Fig.1.7.

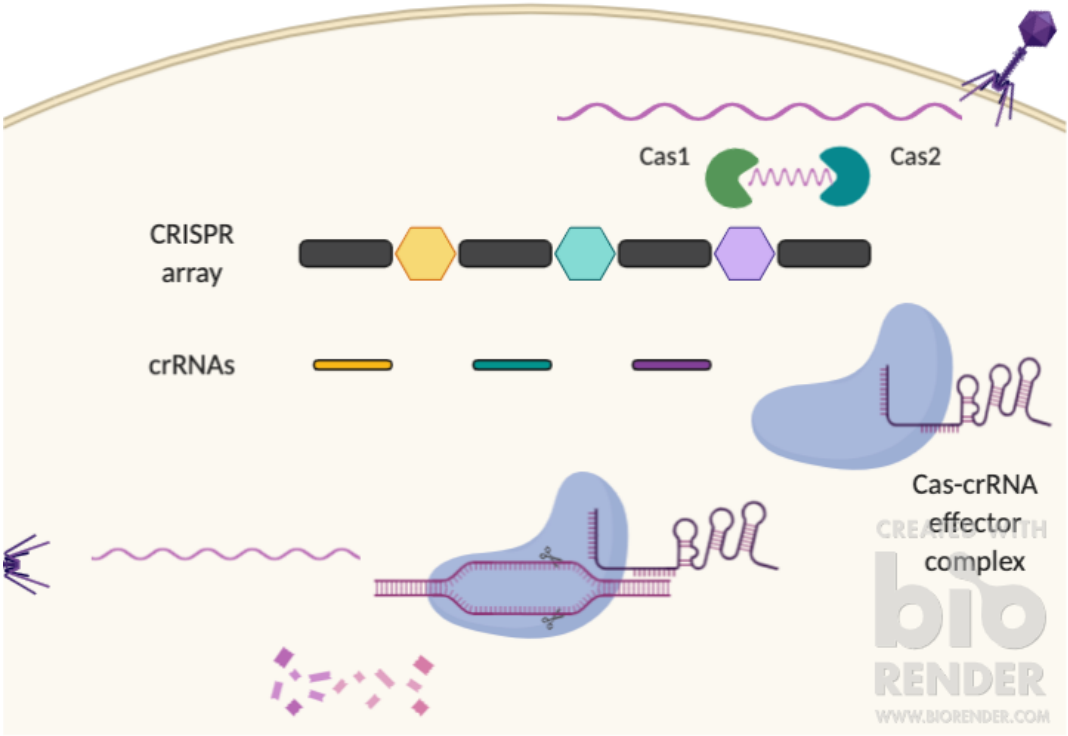


Fig. 1.7 – CRISPR-mediated phage immunity. (Adapted from Jackson *et al.*, 2017). Following phage infection and the insertion of foreign nucleic acid within the host cytoplasm, *Cas1* and *Cas2* proteins incorporate a fragment of the nucleic acid within the CRISPR spacer array. The spacers are transcribed as crRNAs and incorporated within the effector complex, that will use the gRNA sequence to identify homology within the foreign nucleic acid and degrade it.

Briefly, *Cas1* and *Cas2* proteins are responsible for integrating the phage genome within the spacer array of the CRISPR locus. The spacer array is then transcribed into RNA, though the processing of these CRISPR-RNAs (crRNA) is specific to the system subtype. crRNAs are integrated in the effector complex (made up of a single protein for class II CRISPR systems) and guides the effector to recognise the phage foreign nucleic acid thanks to base-pair homology. Upon recognition, the foreign genetic material is cleaved by the nuclease(s).

To avoid cleavage of their own genome, prokaryotes rely on the PAM sequence (Proto-spacer Adjacent Motifs). This is a 2-3bp motif that is located right adjacent to proto-spacers in the foreign nucleic acids injected by phages. The PAM sequences

recognised are specific for the CRISPR/*Cas* system. The effector proteins, once in complex with the crRNA, will only cleave DNA if the PAM is present (Mojica *et al.*, 2009). This is particularly relevant for the applications of CRISPR/*Cas9* to mammalian genome engineering, discussed below.

Fig.1.8 below zooms in on the expression and RNA processing step of type 2 CRISPR systems, focusing on CRISPR-*Cas9*. The processing of the polycistronic pre-crRNA relies on a trans-activating tracrRNA that recognises the repeat sequence in the crRNA and recruits RNAse III to allow cleavage before the crRNA is loaded onto *Cas9*.

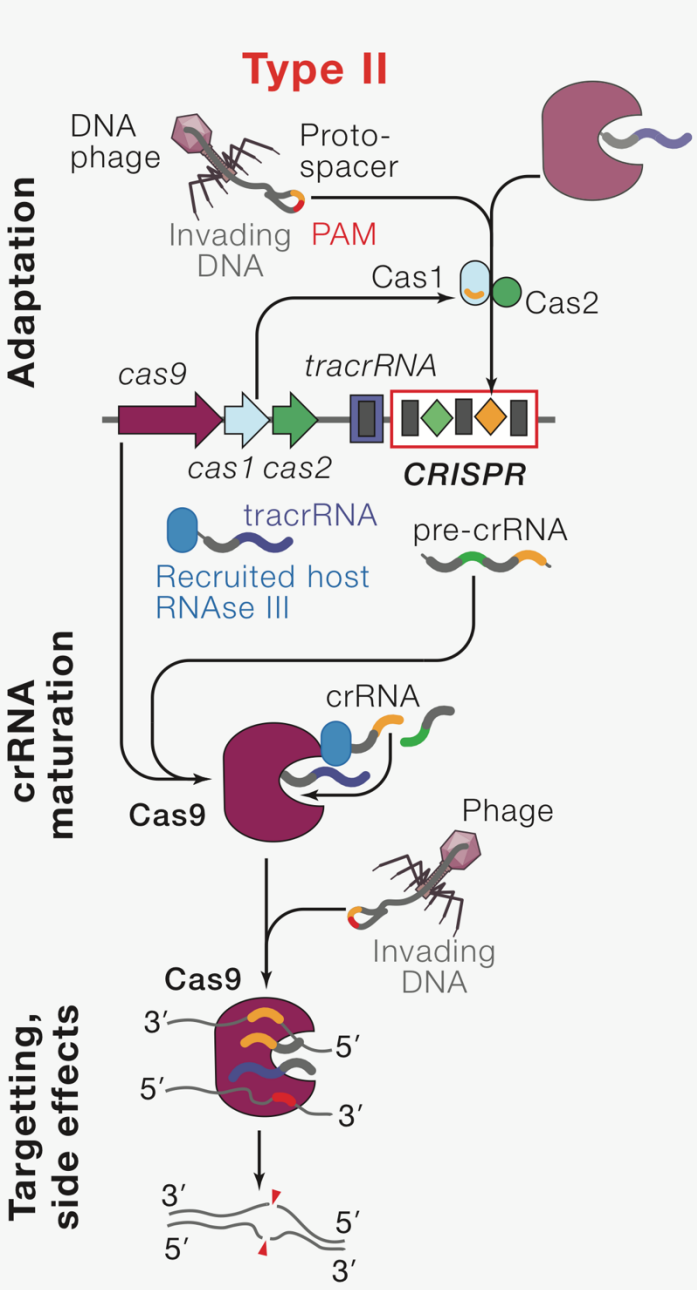


Fig. 1.8 – Type 2 CRISPR-Cas9 specific mediated immunity to phage infections. (Adapted from Makarova, Zhang and Koonin, 2017). In the adaptation step, *Cas1* and *Cas2* proteins integrate proto-spacer sequences from the invading DNA into the CRISPR spacer array. The non-coding trans-activating tracrRNA recognises the repeat sequences in the polycistronic pre-crRNA and recruits the bacterial RNAse III to cleave it, transitioning into the crRNA maturation step. The crRNA is loaded onto *Cas9* while still in complex with the tracrRNA. The tracrRNA is required for crRNA-guided *Cas9* cleavage activity. Upon phage re-infection, the invading DNA is detected by the *Cas9*-tracrRNA-crRNA complex and is degraded by *Cas9* (Jinek *et al.*, 2012).

1.5.4 Translating CRISPR/*Cas9* to genome engineering

The first attempt at using CRISPR-*Cas9* as a cross-species tool occurred in 2011, when Sapranauskas and colleagues transformed the CRISPR array and *Cas9* from *Streptococcus thermophilus* into *E.coli* and demonstrated how it confers it immunity to plasmid transformation and lambda phage infection (Sapranauskas *et al.*, 2011). A year later, *Cas9* from *Streptococcus pyogenes* was purified and shown to cleave target DNA *in vitro* when guided by crRNA (Jinek *et al.*, 2012). The same work showed how the tracrRNA and crRNA could be fused into a chimaeric single transcript, referred to as single gRNA (sgRNA) that allows *Cas9* to cleave any dsDNA that contains the PAM sequence -NGG adjacent to the crRNA homologous region. The authors suggested how this system, requiring expression of only two standard proteins (*Cas9* and bacterial RNAse III) and a single RNA transcript, was advantageous compared to ZFs and TALEs as it was easier and cheaper to customise thanks to its relying on DNA design and synthesis rather than protein engineering.

Again, only one year later, *SpCas9*-mediated genome engineering was achieved in human and mouse cells (Cong *et al.*, 2013; Mali *et al.*, 2013).

Cong and colleagues, from the lab of Prof. Feng Zhang, showed that mammalian RNAses were able to process the pre-crRNA without the need for exogenous bacterial RNAse III to be introduced, which made the system even simpler. After codon-optimisation for mammalian expression of *SpCas9*, and the addition of NLS to enhance *Cas9* localisation to the nucleus, they optimised a strategy consisting of a single expression vector driving *Cas9* expression thanks to the EF1a promoter, and sgRNA expression thanks to the U6 promoter. This construct was transfected and shown to be functional in both mouse and human cells, and it allowed multiplexed genome engineering by providing the cells with a crRNA array targeting independent loci.

Mali and colleagues, from the lab of Prof. George Church, also relied on U6 promoter for the expression of the sgRNA, but expressed codon-optimised, NLS-containing *SpCas9* under the control of the CMV promoter. By transfecting the constructs in

human 293T cells, they showed that they can achieve template-mediated homologous recombination (HR) following *Cas9*-mediated DBS as early as 20h post transfection – as assessed by the expression of a GFP reporter engineered within the hAAVS1 locus. By targeting the ubiquitously expressed PPP1R12C gene in three different human lines using two independent gRNAs, they showed that *Cas9*'s ability to induce NHEJ can be quite variable and depend on both cell line and gRNA used (efficiencies ranging from 10% to 25% for the two separate gRNAs in HEK cells, 13% and 38% in K562 leukaemia cells and 2% and 4% in PGP1 iPS cells). The only reported technical restriction they claim is the requirement for the -NGG PAM sequence, which should represent only a minor limitation as it should occur, considering both DNA strands, every 8bp on average.

As reviewed by Almendros and colleagues (Almendros, Kieper & Brouns, 2019), *SpCas9* is still the most used Cas effector protein for mammalian genome editing, though extensive efforts have been dedicated to optimising its efficiency and to minimise its off-target activity. This is particularly important for therapeutic applications, where off-target effects can't be controlled for by using multiple gRNAs against the same gene (which is sufficient for most research applications) but can have severe consequences on biologicals and/or edited embryos.
A few strategies to address this involved improving gRNA site selection by computationally predicting, identifying and minimising off-target sites. Moreover, *Cas9* mutagenesis of nuclease domains converting them into nickase domains and inducing genome engineering using double gRNAs targeting opposite strands allowed more precise editing (nicks are repaired with much higher fidelity than DBS) (reviewed in Tycko, Myer and Hsu, 2016). Engineering an improved Rec3 domain to improve *Cas9* proofreading activity was also shown to enhance *Cas9* specificity (Chen *et al.*, 2017).

The latest efforts in CRISPR-mediated genome engineering have focused on finding ever smaller effector proteins that would allow easier packaging and delivery through

lentiviruses for instance, that at the same time have non-restrictive PAM sequences to allow easy gRNA site selection. Of note, *Nme2Cas9* from the meningococcus *Neisseria meningitidis* coding sequence is only 3.2kb – compared to the 4.4kb of *SpCas9*, has a permissive PAM sequence (N_4CC) and has been shown to efficiently edit human cell genomes when delivered through lentiviral constructs (Edraki *et al.*, 2019).

The possibilities that CRISPR-*Cas9* has opened are extremely broad, not only focusing on the nuclease activity of the protein, but on the easily customisable targeting properties of its DNA-binding domain. Fig.1.9, taken from a review paper from the Zhang lab from 2014, highlights a few of the potential applications for *Cas9* described soon after its application to mammalian cells had been published. Lentiviral transduction that can reach extremely high efficiencies has facilitated sgRNA delivery and CRISPR/Cas9 applications directly within cells and tissues. DNA labelling and inducible regulation are also mentioned as prominent areas of potential *Cas9* application.

Particularly interesting is the use of catalytically dead *Cas9* (d*Cas9*) bound to an activator domain to induce gene expression. It was first achieved in human cells by tagging d*Cas9* to the activator domain VP64 and targeted to the promoters of 8 different genes. Gene expression induction was verified by RT-qPCR, which showed that the highest expression levels were achieved by synergistic effects of targeting up to 4 sgRNAs to the promoter region of each gene (Perez-Pinera *et al.*, 2013). When it was first established, it was proposed as an alternative way to induce gene expression (compared to more traditional plasmid or lentiviral delivery of the cDNA under an exogenous promoter) and it set out to be a rapid and efficient way to activate multiple genes at once by delivery of a single protein and an array of sgRNAs. In 2016, this system was used to efficiently generate iNs from MEFs by targeting the promoters of *Ascl1*, *Brn2* and *Myt1l* with a combination of 4 sgRNAs per promoter, and introducing d*Cas9*-VP64 expression by lentiviral delivery (Black *et al.*, 2016).

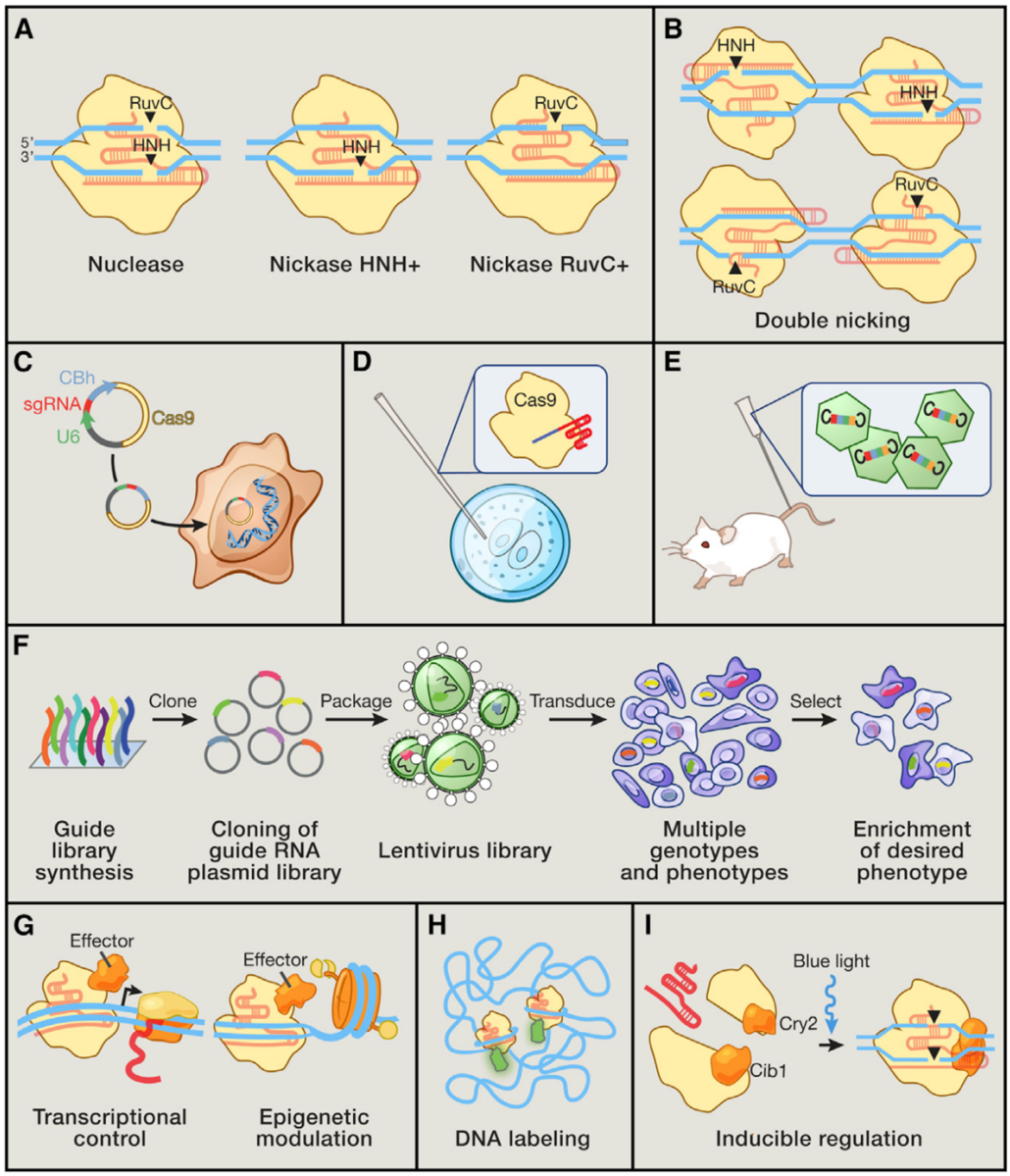


Fig. 1.9 – Cas9 potential applications. (Taken from Hsu, Lander and Zhang, 2014). **(a)** *Cas9* RuvC and HNH domains confer the protein its nuclease activity. By selectively mutating one or the other, *Cas9* can be converted into a strand-selective nickase. **(b)** By mutating both domains, *Cas9* can be converted into a double nickase that can reproduce DBS by appropriately spacing the target site of each nickase. This increases the specificity of DBS as they rely on two independent gRNAs. **(c)** *Cas9* and sgRNA within the same expression plasmids can be introduced as a single construct within target cells. **(d)** Purified *Cas9* protein and sgRNA *in vitro* transcribed molecules can be directly introduced into zygotes to allow fast generation of transgenic mice models. **(e)** Lentiviral delivery of *Cas9* and sgRNA allows efficient introduction of the components into somatic cells and tissues. **(f)** gRNA libraries can

be cloned as a pool into expression plasmids and transduced within cells (together with *Cas9*, or within cells constitutively expressing *Cas9*), that are then exposed to selective pressure to identify the effect of CRISPR-*Cas9*-induced genetic KO on the desired selected-for phenotype. This approach, applied to the conversion between MEFs and induced neurons, is the focus of this thesis – more details are given at the end of this chapter in the "Thesis Aims" section. **(g)** d*Cas9* can be coupled to effector domains to selectively activate or suppress gene expression. This can be used to understand the role of promoters or enhancers by designing sgRNAs against those loci. **(h)** Tagging d*Cas9* with fluorescent reporters can be used to follow chromatin dynamics and track the behaviour of specific loci during chromosome rearrangement, or nuclear lamina association, or the assembly of transcription factories and so on. **(i)** Inducible genetic regulation can be achieved by splitting up *Cas9* into heterodimer domains and relying on chemical or optical induction to initiate protein assembly and inducing its activity.

1.5.5 CRISPR/*Cas9* in genetic screening

Building on the work of the Zhang and Church labs, in 2014 3 seminal papers described the first CRISPR/*Cas9*-mediated loss-of-function genetic screenings in mammalian cells (Koike-Yusa *et al.*, 2014; Wang *et al.*, 2014; Shalem *et al.*, 2014).
In all three papers gRNAs were delivered as a lentiviral pool – and expressed under the control of the U6 promoter. The gRNA sequences themselves, integrated within the host genome thanks to the lentiviral delivery, would serve as barcodes to allow NGS to count the number of cells selected for or against a particular gRNA.
Wang and colleagues established an engineered stable line with constitutive expression of *Cas9*. This resulted in circa 97% of on-target indels, with the majority of DBS resulting in small deletions (<20bp) leading to non-functional proteins, and minimal off-target editing (as assessed by sequencing of predicted off-target sites).
gRNA design relied on -NGG, targeting constitutive coding exons, as early as possible within the coding sequence to introduce premature stop codons and trigger nonsense-mediated mRNA decay, and filtered for the lowest amount of predicted off-target sites within the human genome. Multiple sgRNAs were designed per gene

– to account for different editing abilities of gRNAs and to ensure that, in the analysis step, a false-negative or false-positive gRNA read count could be controlled-for by looking at read counts of gRNAs targeting the same gene.
They also discuss the highly variable efficacies of gRNAs by looking at multiple gRNAs targeting the same genes. GC content, targeted exon position and targeting the transcribed vs non-transcribed strands all seemed to be factors involved in gRNA efficacies. gRNA purine/pyrimidine content in the 3' end was also a strong predictor for efficacies. Together, these data were used to generate a genome-wide sgRNA targeting library of predicted high-efficiency.

Shalem and colleagues instead delivered *Cas9*, the sgRNA and a puromycin selection marker within a single lentiviral construct. This makes it easier to perform genome engineering and genetic screens in any cell without the need to generate a *Cas9*-expressing line first, but is susceptible to more variability as the locus of lentiviral integration can potentially affect *Cas9* expression and lead to heterogeneous levels in the examined cells, with consequences on editing efficiencies.
The library they designed, called GeCKO (for Genome-scale CRISPR-*Cas9* knock-out), targets 5' constitutive exons with, on average, 4 gRNAs per gene, designed using similar considerations in terms of off-target predictions as the Wang paper.
They used this library to perform a genome-wide CRISPR-*Cas9* screen looking at identifying genes involved in PLX drug resistance in A375 cells and compared the result with a RNAi screen for the same resistance in the same cell line, showing better consistency in gRNA behaviour compared to shRNAs. They also validated up to 5 gRNAs for top-ranking genes using deep sequencing and verified that 23/25 sgRNAs resulted in >95% allele modification.

Whereas the works by Wang and Shalem and their colleagues focused on human cells and generated genome-wide gRNAs libraries targeting the human genome, Koike-Yusa and colleagues performed a similar genome-wide screen on mouse embryonic stem cells instead. Similarly to Wang and colleagues, they produced a mESC line

constitutively expressing *Cas9* under the control of the EF1a promoter, and transfected cells with sgRNAs, showing a higher knock-out efficiency from the *Cas9*-expressing line, compared to wt cells that received *Cas9* expression via transfection of a separate plasmid.
For screening purposes, the gRNAs were delivered as a pool within a lentiviral construct driving gRNA expression under the U6 promoter, as the previous two studies. The library was designed with 5 gRNAs per protein-coding genes in the mouse genome.

All three studies show evidence of the success of CRISPR/*Cas9* loss-of-function positive selection screens, by exposing the cells to drugs or toxins to identify genes involved in susceptibility to those agents. Cells in which susceptibility genes have been knocked-out after being targeted by gRNAs will survive. This set-up exposes the cells to strong selective pressure and will result in low background, as cells that fail to acquire a resistance phenotype by genetic KO will die off, whereas surviving cells will proliferate and biologically "amplify" the representation of the gRNAs that target susceptibility genes within the surviving cell population.

The studies by Shalem et al. and Wang et al. also show evidence of the success of loss-of-function negative selection screens. This type of screen looks for essential genes involved in a process by exposing cells to selective pressure and looking for gRNAs that are underrepresented or depleted within the final cell population. The two studies wanted to investigate genes essential for cell viability: they transduced the cells with gRNA libraries and allowed them to proliferate for several cell doublings. Cells that received gRNAs targeting genes essential for survival will die off. By looking for underrepresented gRNAs in the pool of surviving cells, genes involved in cell viability can be identified. Compared to positive selection screens, this approach is likely to reveal a broader range of genes – some genes might be completely essential, while the knock-out of others might just slow down proliferation. This was in fact the case for the Shalem and Wang papers, and their

analysis focused on the most strongly depleted gRNAs to identify essential genes – revealing in fact genes involved in DNA replication, gene transcription and translation.

The analysis of the NGS dataset itself had relied on customised algorithms or, in the case of Shalem et al., the RIGER algorithm previously developed to rank enrichment of shRNAs from RNAi screens (Luo *et al.*, 2008). In 2014, the MAGeCK algorithm (Model-based Analysis of Genome-wide CRISPR/*Cas9* Knockout) was developed specifically to analyse NGS screening datasets containing gRNA read counts from CRISPR-*Cas9* screens, overcoming the limitations of previous algorithms focused on oligonucleotide barcodes from microarray data, different sequencing depths, different mapping of sequences to genes and pathways, and taking into consideration for instance the number of gRNAs targeting each gene.

These seminal studies highlighted how translatable CRISPR-*Cas9* technology was, and how the right tools had allowed with relative ease to perform genome-wide screens on mammalian cells with more hits, and more robust hits, compared to traditional RNAi screens. To summarise a few key points:

1. they showed how *Cas9* integration within cell lines allows strong editing efficiencies and full genetic KO in mammalian diploid cells within one day of gRNA integration
2. they showed that lentiviral delivery of gRNAs, compared to transfection, allows robust gRNA expression
3. they showed that a single, lentivirally-delivered copy of the gRNA expression construct is enough to induce efficient DBS
4. they established genome-wide sgRNA libraries for human and mouse cells – a great resource for all labs working on those species. The Yusa lab in 2016 improved on their mouse genome-wide gRNA library by perfecting sgRNA design and minimising off-targets (Tzelepis *et al.*, 2016). This library was used

in this project to select gRNAs against the chosen gene subset – more information is provided in Chapter 3.

Since then, many labs have used CRISPR-*Cas9* screens to investigate a myriad of different biological processes, as reviewed by several publications (Kweon & Kim, 2018; Joung *et al.*, 2017; Doench, 2018). Screens have not only focused on loss-of-function phenotypes, but thanks to all the mutagenesis on *Cas9* and the optimisation of gRNA targeting, they have been used for gene activation or repression, DNA methylation or de-methylation, histone modifications and even DNA base editing.
In the context of *in vitro* derived neurons, a recent CRISPR activation screen targeted circa 2500 predicted transcription factors and DNA binding factors with a total of 55'000 gRNAs to identify genes that, when overexpressed, would convert mouse ESCs into neurons (Liu *et al.*, 2018).
They then overexpressed their top 14 hits in MEFs to check whether they could convert them into iNs, revealing that *Ngn1*, and a combination of other factors and BAM factors, can convert to some extent fibroblasts into iNs.
This approach is particularly interesting as it highlights a quite stringent library selection (targeting only 2500 genes) that is based on biased gene predicted function (DNA binding and transcription activation), but not gene expression.

Overall, the CRISPR-*Cas9* screening technologies represent a powerful tool to investigate the function of multiple genes at once in the context of a specific biological process. I decided to apply this technology to identify and investigate which genes are involved in the *in vitro* conversion between MEFs and iNs.

1.6 Aims of this thesis

The aim of this thesis was to identify genes involved in the direct conversion between fibroblasts and neurons and use this information to make the process more

deterministic, thus increasing the efficiency of the conversion and neuronal purity of the culture.

To achieve this, I designed a CRISPR/*Cas9*-mediated screening strategy to investigate gene roles by looking at loss-of-function effects. The experimental design is outlined in Fig.1.10 below. *Cas9*-expressing MEFs transduced with the BAM factors to allow conversion into iNs were also transduced with a pooled lentiviral gRNA library targeting a subset of genes that, based on their gene expression, are likely to be involved in the process.

Thanks to the lentiviral delivery, gRNAs transduced within MEFs allow gene KO while being integrated within the host genome and propagated as cells grow and convert.

After 15 days, the neurons were isolated from the rest of the cell types within the culture, and, using Next Generation Sequencing (NGS), the distribution of the gRNAs integrated within the pool of neurons was analysed (how to isolate, using flow cytometry, the iNs from the heterogeneous conditions is also addressed in Chapter 3).

The distribution of gRNAs within the neuronal genomes should allow the identification of essential genes or facilitators of the conversion, as well as potential roadblocks that hinder the conversion.

The former should be identified by a depletion of their targeting gRNAs, since cells that received a gRNA against an essential conversion gene would have been less likely to become neurons, and thus those gRNAs would be present in fewer copies in the sequencing data.

Vice versa, the latter should be identified by an enrichment of their targeting gRNAs, since MEFs that received a gRNA against a conversion roadblock would have higher chances of converting into a iN, and thus those gRNAs would be present in more copies in the sequencing data. The MAGeCK algorithm was used to combine the behaviour of individual gRNAs targeting the same gene and identifying genes that are overall enriched or depleted, and thus potential roadblocks or enhancers of the conversion.

Chapter 3 addresses the system set-up: how to isolate the neurons, and how the library of genes to screen against was designed and prepared.

Chapter 4 addresses the screen process and its validation, and discusses the role of *Sf3a1* in the conversion.

Chapter 5 addresses the characterisation of 20 individual genes selected based on their known role in the conversion and highlights the role of *Stxbp1* in iN survival.

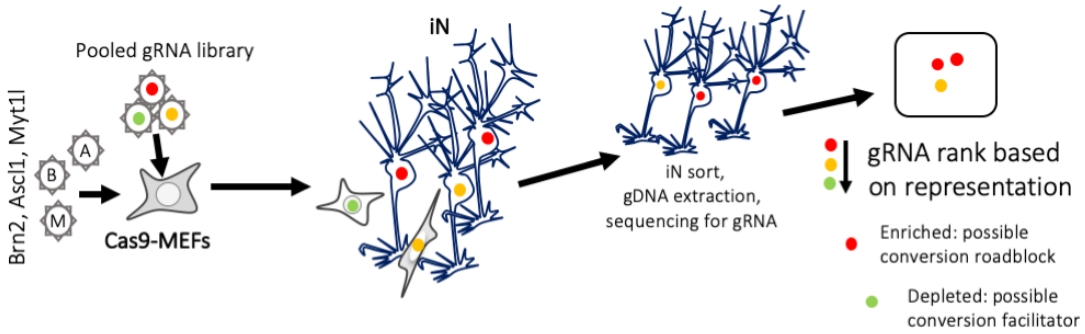


Fig. 1.10 – Screen experimental design. *Cas9*-MEFs transduced with the BAM factors were converted into iNs. A pooled gRNA lentiviral library transduced together with the factors allowed gRNA integration within the host genome and *Cas9*-mediated gene KO. Neurons after 15 days from the start of the conversion were isolated and the gRNAs integrated within their genomes were sequenced and their distribution quantified. In red is represented a gRNA targeting a potential conversion roadblock gene. Upon KO, the MEFs that receive this gRNA would have a better chance of becoming neurons, thus this gRNA would be enriched within the neuronal genomes.

In green is represented a gRNA targeting a potential essential gene for the conversion. This gRNA would prevent the MEFs that receive it to convert into iNs, thus it would not be found in the neuronal pool and would be underrepresented in the sequencing data from neuronal genomes.

In yellow is represented a gRNA targeting a gene that has no effect on the conversion, and whose KO would not impact the conversion. Thus, this gRNA would be equally represented compared to the original gRNA distribution in the lentiviral pool.

1.7 A note on *in vitro* neuronal identity

Overall, neuronal identity is defined by 4 major, equally important characteristics.
Firstly, morphology: neurons are recognizable by the presence and defined length of dendrites and a single neurite sprouting from the cell body. This feature was used by the first iN publication (Vierbuchen et al., 2010) to identify and quantify iNs via imaging. They defined a neuronal morphology as a cell having a neurite whose length is at least 3 times the diameter of the cell body.
Of course neuronal morphology changes and increases in complexity as the cell matures and creates connections with its neighbor. The criterion used by the first publication has become widely used as a minimal morphological criterion to identify an *in vitro* induced neuron.
In this thesis, I've used this criterion to identify and quantify iNs in my culture system.

Secondly, the expression of a subset of specific markers that are exclusively expressed by cells with neuronal morphology in the culture and can distinguish iNs from fibroblast, myoblasts or neural but non-neuronal cells. Among the most commonly used, and first used by the Wernig lab in 2010, are *Tubb3, Map2, NeuN, Tau, Synapsin.* Expression of these genes marks any newly converted iN regardless of the expression of subtype-specific markers.
In this thesis, I've used these markers to identify and quantify iNs in my culture system. I haven't addressed the subtype identity and the heterogenicity of neuronal phenotypes in the culture. It's important to highlight this as my results are potentially impacted by the different neuronal subtypes but I don't have a way of examining this further with the tools and platforms discussed here.

Thirdly, electrophysiological properties of the cellular membrane are a defining feature of a neuron. A neuron has distinctive resting membrane potential and has to be able to fire an action potential when the membrane is stimulated with a voltage above a certain threshold.

This feature has been addressed by most of the publications on iNs. The original paper reported electrophysiological activity of the iNs upon co-culture of the cells with astrocytes or ex-vivo cortical neurons from day 13 onwards of the conversion. It is important to note that before the iNs reach this stage of maturation, they should be identified as immature iNs – as lacking functional evidence of neuronal electrical characteristic behaviour.

Electrophysiological studies are not present in this thesis, as they were beyond its scope. It has to be mentioned hence that it is likely my assessment of iNs, that relies on morphology and marker expression alone, includes a heterogeneous mixture of cells of electrical immature nature.

Lastly, a neuron should be defined by its subtype. As I discussed in section 1.2.2, extensive work has been done to generate pure populations of iNs of a specific subtype that express markers relative to the neurotransmitter and pathway the neurons produce and partake in. The original 2010 publication on iNs reported cells reprogrammed with Ascl1, Brn2 and Myt1l as having heterogeneous subtype identity. Presence of GABAergic neurons, glutamatergic neurons was assessed and quantified both via immunofluorescence and electrophysiological responses. Dopaminergic neurons seemed to be absent. Several of the cells identified as iNs lacked the expression of subtype-specific markers. It's important to remember this when defining an induced neuron and using these cells for disease modelling or drug testing.

In this thesis, I did not address the neuronal subtype identity of the iNs produced as it went beyond the scope of the study. It's then important to note that the cells I define as iNs or pan-iNs are marked by morphology and expression of canonical neuronal markers, but are likely to be a mix of cells with specified subtype identity and cells lacking one.

Chapter 2 Materials and methods

2.1 Cell culture

2.1.0 General notes

All cell cultures were maintained at 37ºC, 5% CO_2 in normoxic conditions. All media and reagents, unless where specifically stated, were warmed up to 37ºC in a water bath before being added to cells.

Cell centrifugations were performed on a tabletop centrifuge for 3 minutes at 1300rpm.

Cell thawing was performed by warming up the frozen vials at 37ºC in a water bath for 1 minute before adding the solution of thawing cells to 10ml of warm medium. Cells were then spun, resuspended in appropriate medium and plated at the appropriate density in culture flasks, dishes or wells.

Cell freezing was performed by harvesting the cells using 0.25% Trypsin-EDTA, washing in 10% FCS medium and resuspending spun pellets in FCS + 10% DMSO. Cells were frozen down in Mr. Frosty™ Freezing Container (ThermoScientific) at -80ºC for 24h before being moved to LN_2 long term storage.

2.1.1 Cell culture reagents

HEK medium

- Glasgow Minimal Essential Medium (GMEM, Sigma G5154)
- Fetal Calf Serum (FCS) (10%) (Life Technologies)
- Non-essential amino acids (1X, Gibco 11140-035)
- L-Glutamine (2 mM, Invitrogen)
- Penicillin/Streptomycin (100 units/100 µg, Sigma P4333)

MEF medium

- Glasgow Minimal Essential Medium (GMEM, Sigma G5154)
- Fetal Calf Serum (FCS) (10%) (Life Technologies)
- Non-essential amino acids (1X, Gibco 11140-035)
- L-Glutamine (2 mM, Invitrogen)
- Sodium pyruvate (Invitrogen)
- Penicillin/Streptomycin (100 units/100 µg, Sigma P4333)
- β-mercaptoethanol (ThermoFisher Scientific, 31350010)

Conversion medium (CM)

- Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham (Sigma D8437)
- Glucose (Sigma G8644)
- MEM NEAA 100x (Gibco 11140-035)
- Pen-Strep (Gibco 15140-122)
- BSA Solution 0.012% (Gibco 15260-037)
- β-mercaptoethanol 50mM (Gibco 31350-010 (20ml))
- B27 Supplement 50x (Life Technologies 17504-044)
- N2 Supplement 100x (Life Technologies 17502-048)
- Laminin 1µg/ml (Sigma L2020-1mg/ml)

Freezing solution

- Fetal Calf Serum (FCS) (Life Technologies) – 90%
- DMSO (WR International) – 10%

Other reagents

- Trypsin-EDTA (0.25%, Gibco 15090-046)

- PBS (Sigma, D8537)

2.1.2 Cell lines used

- Cas9 MEFs – isolated as described below (paragraph 2.1.3). MEFs were isolated from mice crosses between wild type animals and animals from the C57BL/6 KYEC Rosa26$^{+/sNLSCas9}$ line, obtained from the lab of Dr. Kosuke Yusa. This line is heterozygous for the Rosa26 locus as one locus harbours the coding sequence for *SpCas9* containing a single nuclear localisation signal (sNLSCas9). A homozygous mouse line containing two copies of *SpCas9* became available only at the end of the project and thus was not used for the work presented in this thesis.

- HEK293T – lentivirus packaging line

2.1.3 MEF isolation

Mice crosses were set-up between female WT animals and male C57BL/6 KYEC Rosa26$^{+/sNLSCas9}$ animals. Embryos were isolated from culled mothers at E14.5 and kept in cold PBS supplemented with 2% Pen/Strep (Sigma P4333). Under a dissection microscope heads, internal organs, spine and tail were removed. A portion of fetal liver was kept aside for genotyping. The remaining tissue was incubated in 0.5ml 0.25% Trypsin / 2 mM EDTA for 15 minutes at 37ºC, and then mechanically disassociated by passing it through an 18-gauge needle. The suspension was added to 10ml of MEF medium to allow trypsin quenching. Each embryo was handled individually, and the final solution of 10.5ml was plated into a 10cm^3 dish. The next day the medium was changed to remove residual trypsin. Cells were allowed to expand for 48h from seeding, then pooled based on genotyping results and frozen. Genotyping results consistently confirmed 50% Mendelian inheritance of the Rosa26$^{+/sNLSCas9}$ locus.

2.1.4 MEF genotyping

DNA extraction solutions:

1) Lysis buffer: 25mM NaOH / 0.2 mM EDTA
2) Neutralisation buffer: 40mM Tris-HCl (pH 5.0)

The portion of fetal liver kept per embryo was added to 30µl of lysis buffer and incubated at 95ºC for 20 minutes. After rapid cooling to 4ºC, 30µl of neutralisation buffer were added to the samples, followed by thorough vortexing. 1µl of this lysate was used for the genotyping PCR in a 15µl final volume as shown in Table 2.1. Table 2.2 explains the PCR running conditions. Amplicons were then analysed on a 2% agarose gel. Primer sequences are in Table 2.3.

Cell lysate	1 ul
2.5 mM dNTP	1.2 ul
DreamTaq Buffer 10X	1.5 ul
DreamTaq Polymerase (Thermo Scientific)	0.075 ul
20 µM F primer	0.375 ul
20 µM R primer	0.375 ul
dH2O	10.475 ul

Table 2.1 – Genotyping PCR solution mix

95C	3'	
95C	10s	x20
60C	10s	
72C	30s	
72C	5'	
4C	cooling	

Table 2.2 – PCR running conditions

Cas9-U2	CTCTCCCAAAGTCGCTCTGA	Rosa26 F	
Cas9-L2	ACCCCAGATGACTACCTATCCT	Rosa26 Wt R	395bp amplicon
Cas9-L3	GAAAGACCGCGAAGAGTTTGTC	Rosa26 Cas9 R	317bp amplicon

Table 2.3 – Cas9 genotyping primer sequences.

2.1.5 MEF freezing

Based on genotyping results, cells derived from embryos carrying the Cas9 transgene were pooled together upon harvesting. Cells were lifted by adding 1ml 0.25% Trypsin / 2 mM EDTA to the dish. After 5 minutes incubation at 37ºC, the solution was quenched with 9ml MEF medium. Cells were centrifuged for 3 minutes at 1300rpm, counted and then aliquoted as $2.5x10^6$ cells per cryovial in freezing medium and stored in LN_2 long term.

2.2 Lentiviral production

The lentiviral particles used in this work were produced using a 2nd generation packaging system, where the *gag*, *pol* and *rev* packaging sequences are expressed from the psPAX2 vector, while the encapsulation sequence from the VSV-G *env* gene is expressed from the pMD2.G vector. The sequence of the construct of interest to transduce is contained within flanking self-inactivating long terminal repeat (LTR) regions. Source of vectors can be found in table 2.18.

2.2.1 HEK culture

HEK from communal lab batches frozen at as $5x10^6$ cells per vial were thawed, seeded in a t150 flask and allowed to grow for 2-3 days.

HEKs were routinely passaged upon reaching 80% confluency by lifting them with 0.25% Trypsin / 2 mM EDTA and re-seeding them to a density of 1:5 to 1:10 of the original one, every 3-4 days.

Early passage batches were expanded and frozen at a density of $5x10^6$ cells per cryovial in freezing medium and kept in LN_2.

2.2.2 HEK seeding and transfection

One day before transfection, HEKs were lifted, counted, and seeded at a density of 2 million cells per $10cm^2$ dish. The next day, cells were transfected with the transfer vector plus psPAX2 and pMD2.G using the calcium phosphate protocol. Plasmids and water were mixed to a final volume of 437μl, then 63μl of 2M calcium chloride were added to the mix (126mM final concentration). The solution was quickly vortexed. While vortexing, 500μl of HBS (8 g NaCl, 0.2 g Na2HPO4-7H2O, 6.5 g HEPES) were added dropwise to allow the formation of homogeneously sized precipitate. After 10-minute incubation at room temperature, the solution was added dropwise to the HEK dish, then shaken to ensure even distribution of the transfection mix.

The next day, the HEK supernatant was removed and changed for 7ml GMEM.

2.2.3 Virus harvesting and freezing

Viruses were harvested at 48h and 72h post-transfection. After the first harvesting, 7ml GMEM were added to the cells to allow for the second collection. The supernatants were filtered through 0.22μm PES filters and stored at 4ºC up to 24h before pooling, aliquoting, and transported in dry ice to long-term storage at -80ºC.

2.2.4 Virus concentration

For the tetO-Ascl1, Brn2, Myt1l and hUb-rtTA constructs, lentiviral particles from supernatant were concentrated to allow co-transduction of the 4 factors within cells

in small volume amounts. Filtered fresh supernatants were mixed with Lenti-X™ Concentrator (TaKaRa #631232) in a volume ratio of 3:1 and incubated overnight at 4ºC. The next day, the solutions were centrifuged for 2h at 4ºC at 1500rpm. This allows lentiviral particles bound by the Lenti-X concentrator polymer to be spun down. The supernatant was then discarded and 1/100th of the original supernatant volume of GMEM was added and incubated overnight with the lentiviral pellets at 4ºC. The next day the pellets were gently resuspended, aliquoted and transported in dry ice to -80ºC for long-term storage.

2.2.5 Virus titration

Lentiviral titre was assessed by transduction into MEFs and analysis -either IF or flow cytometry depending on the construct at 48h.

Both processes gave an accurate estimate of a functional titre – i.e. the volume of lentiviral solution required to transduce a specified percentage of MEFs, rather than an estimate of actual lentiviral particles in the supernatant.

Predicted MOI was estimated and adjusted during transductions based on % of transduced cells during functional titrations, following the mathematical formula for probability and MOI. For instance, MOI 1 corresponded to a transduction efficiency of 60%.

24h before transduction, MEFs were seeded in 12wells at a density of 50'000 cells per well, the same density as a conversion experiment. On the day of transduction, lentiviral aliquots were thawed by a short incubation at 37ºC, then mixed with GMEM to obtain the desired dilution and Polybrene (Merck-Millipore, TR- 1003-G) at a final concentration of 10 µg/ml.

For sgRNA vectors – which harbour an EF1α-BFP reporter, MEFs were transduced with lentiviruses either undiluted or at a concentration of 1:3 and 1:6 of the original stock. Cells were incubated with lentiviruses for 4h, then the medium was reverted to MEF medium. 48h post-transduction, flow cytometry was used to assess the % of

transduced cells by examining BFP reporter activation. Each sgRNA vector batch was tested individually and against every new MEF batch to minimize experimental variation.

tetO-Ascl1, Brn2, Myt1l were co-transduced with hUb-rtTA vectors in MEFs using 10+10, 5+5 and 2.5+ 2.5 µl of concentrated lentiviral solutions. Cells were incubated with the viral particles for 4h, then the medium was reverted to MEF medium. The next day, doxycycline at a concentration of 2µg/ml was added to induce TF expression. After 48h, cells were assayed by immunofluorescence to estimate the percentage of transduced cells.

Titre of the hSyn1-dsRed construct was estimated by transducing converting cells at day 3. Cells were incubated with the lentiviral particles for 4h, then the medium was reverted to TM. At day 15, the number of iN co-expressing *Tubb3* and *dsRed* was assayed by immunofluorescence.

2.3 Conversion experiments

Every conversion experiment involved thawing with a fresh vial of MEFs, as serial passaging of MEFs induced cell senescence and reduced the quality of conversion.

MEFs were thawed on day -4 before transduction and fed the next day. On day -1 cells were lifted, counted and seeded at a density of 50'000 cells/cm^2 in 1ml of MEF medium per 12 well.

The next day, cells were transduced with lentiviral particles containing cDNA sequences for tetO-Ascl1, tetO-Brn2, tetO-Myt1l and hUb-rtTA at MOI5, resuspended in GMEM supplemented with 10µg/ml Polybrene (Merck, TR-1003-G). Cells were incubated with the transduction mix for 4 hours, then the medium was changed to MEF medium. The next day, cells were given another medium change and provided with MEF medium supplemented with doxycycline (2µg/ml, Sigma D9891).

On day 3, the medium was changed to conversion medium (CM) supplemented with 2µg/ml doxycycline. Medium was changed every 2-3 days after then until the end of the experiment.

2.3.1 Pilot Screening

Cas9-MEFs were seeded at a density of 100'000 cells per 6well as per a standard conversion experiment, one day before transduction. Cells were then transduced with the BAM factors and rtTA at MOI5, plus gRNAs against Ascl1 (#1 and 2), intron1 of Pecam1 and Zeomycin so that 2250 MEFs out of the total 100'000 cells seeded would be infected per gRNA. This mimicked the screen coverage within a single 6well – pilot was performed in duplicate 6wells. As a control, the same gRNA mix was transduced into Cas9-MEFs who did not receive the reprogramming factors. One day post-transduction, doxycycline was added to induce TF expression. After 72h, converting MEFs were transduced with hSyn1-dsRed at MOI5 to mark iNs. At day 15, dsRED+ve neurons were sorted via flow cytometry, genome harvested, and gRNA presence was assessed with qPCR, using Brilliant III Ultra-Fast SYBR PCR (Agilent) as a reaction mix, and run in the LightCycler480. gDNA was also harvested from control wells of non-converting, gRNA transduced MEFs.
A common forward promoter was used for all gRNAs – it anneals to the U6 promoter within the expression construct. The reverse promoter represents the reverse complement of each gRNA sequence. Sequences are reported Table 2.4 below.

Name	**Sequence**
Common U6 F	TACGATACAAGGCTGTTAGAGA
Ascl1 gRNA #1	CCGGGAGCATGTCCCCAAC
Ascl1 gRNA #2	GGTTGACCAACTTGACCCG
Zeomycin	GAGGCGTCCCGGAAGTTCGC
Pecam1	GTCACAAAGAACCCAAGAGG

Table 2.4 – list of gRNA primer sequences used for pilot screening.

2.3.2 Screening

9 million Cas9-MEFs per replicate were seeded in 15x6wells to a density of 100'000 cells per well one day before transduction. Cells were then transduced with BAM factors and rtTA at MOI5, and the lentiviral gRNA library pool at MOI3. 24h post-transduction, doxycycline was added to the media to induce TF expression – this was termed day 0. On day 3, cells were transduced with hSyn1-dsRed, then the medium was switched to neural conversion medium. On day 15, cells were flow-sorted to isolate the dsRED+ve and dsRED-ve populations. A representation of the gating strategy is shown in Fig.2.1 below.

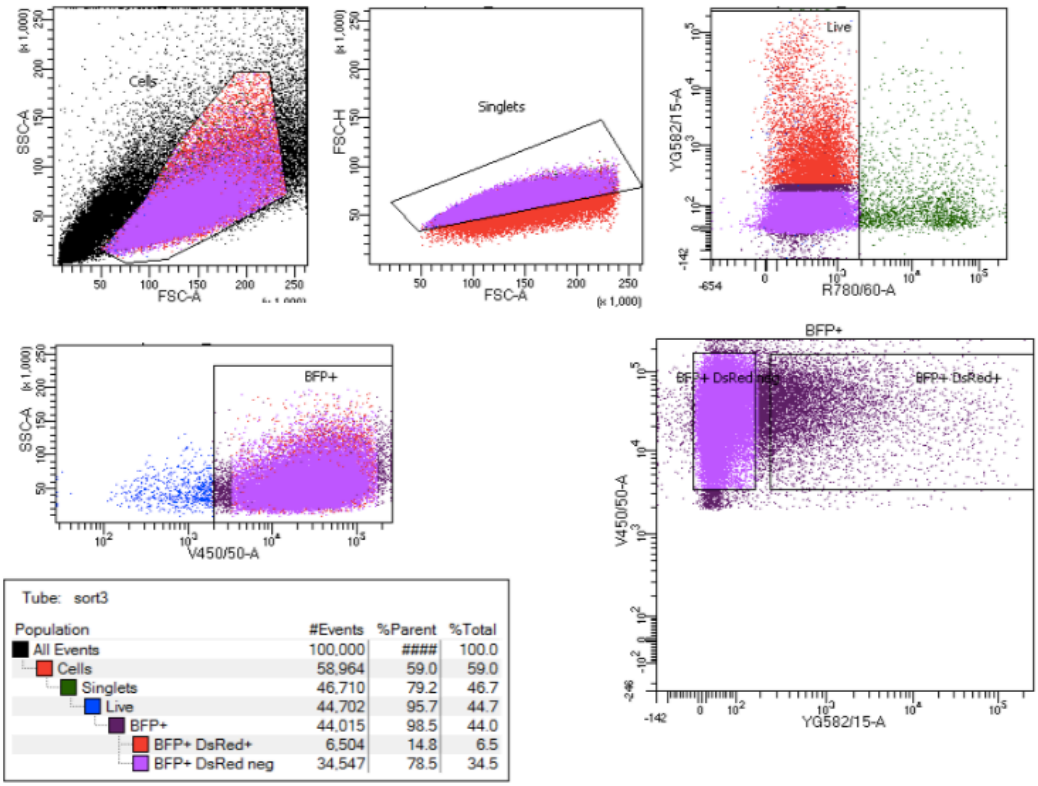


Fig. 2.1 – gating strategy for the screening samples. First, cells are separated from debris and aggregates based on the SSC-A/FSC-A plot. Then, singlets are identified based on FSC-H/A ratio. Dead cells are excluded based on the absence of signal from Draq7 (Abcam), detected on the R780/60 channel. These are then gated on BFP+ve cells (that received and

express the gRNA cassette), detected on the V450/50 channel. Finally, two gates within BFP+ve cells are drawn to sort dsRED-ve and dsRED+ve cells, detected in the YG582/15 channel. Plots are taken from one of the three replicates.

2.4 Flow cytometry and FACS

Flow cytometry materials

- Flow cytometry buffer
    - PBS (Sigma, D8537)
    - Fetal Calf Serum (FCS) (Life Technologies) – 2%
- Live/dead marker Draq7 (ab109202), 1µM final concentration
- Antibodies
    - Rat α-GMP6A (OriGene AM26432AF-N), 1:250
    - Rabbit α-Syp (LSBio aa225-253), 1:1000
    - Mouse α-Snap25 (Abcam ab66066), 1:1000
    - Goat α-Rat IgG AF568 (ThermoFisher A21434), 1:1000
    - Donkey α-Mouse IgG AF568 (ThermoFisher A10037), 1:1000
    - Donkey α-Rabbit IgG AF488 (Abcam ab150073), 1:1000

2.4.1 Cell harvesting and staining for flow cytometry

To prepare cells for flow cytometry, wells were washed once with PBS, then 0.5ml Trypsin-EDTA was added onto the monolayers and incubated at 37ºC for about 5 minutes. MEF medium containing 10% FCS was added to the wells to quench Trypsin (at least 4x the Trypsin volume). Cell suspensions were spun down and washed in PBS, then processed for staining or resuspended directly in FACS buffer containing Draq7 (for live samples).

Staining for GPM6A was performed on live cells: the primary antibody was incubated for 15’ on ice in FACS buffer. After 3 washes, cells were stained with the secondary antibody for 5’ on ice in the dark, again in FACS buffer. Cells were then washed 3

times, resuspended in FACS buffer containing Draq7 and brought to analysis to the FACS machines.

Staining for Syp and Snap25 was performed on fixed cells: cells suspensions were fixed for 10' at room temperature with 4% PFA, then permeabilised for 20' at room temperature with 0.1% Tween-20 in PBS, then blocked for 30' at room temperature in 5% BSA in PBST (0.1% Tween-20 in PBS). Antibody staining was then performed in the same conditions as described above, except 5% BSA was used as a staining solution containing antibodies, and PBS was used to wash and to process cells in the FACS machines.

2.4.2 Flow cytometry analysis and laser settings for fluorochromes

Right before flow cytometry, samples were passed through a 40 µM mesh filter (BD) to avoid machine clogging due to cell clumping. FACS sorting was performed on the BD FACS Aria II by the SCRM flow cytometry team (Dr. Fiona Rossi, Dr. Claire Cryer, Dr. Bindy Heer, Dr. Andrea Corsinotti). Flow cytometry analysis was performed on the BD FACS Fortessa. Compensation was performed manually when required using the built-in DIVA software (BD). Tables 2.5 and 2.6 below summarise the lasers and filters used for flow cytometry in both instruments.

BD FACS Aria II		**Excitation laser**		
		405nm	**561nm**	**640nm**
Filter	**450±50**	**BFP**		
	582±15		**dsRED**	
	780±60			**Draq7**

Table 2.5 – BD Aria II lasers and filters for cell sorting

BD Fortessa		Excitation laser			
		405nm	488nm	561nm	640nm
Filter	450±40	BFP			
	525±50		AF488		
	582±15			dsRED AF568	
	780±60				Draq7

Table 2.6 – BD Fortessa lasers and filters for flow cytometry analysis

2.5 Imaging Cytometry

2.5.1 Immunofluorescence staining

Staining solutions

- Fixing solution: 4% PFA (Sigma-Aldrich) in PBS
- Permeabilizing solution: 0.1% TritonX (ThermoScientific) in PBS
- PBST: 0.1% Tween-20 (Fisher BioReagents) in PBS
- Blocking solution: 5% BSA (Sigma-Aldrich) in PBST

Primary antibodies

- Mouse α-Tubb3 (clone #Tuj-1), (R&D Systems MAB1195), 1:1000
- Rabbit α-Tubb3 (Abcam ab18207), 1:1000
- Mouse α-MAP2 (Sigma-Aldrich M1406), 1:500
- Rabbit α-RFP (Abcam ab62341), 1:250
- Rat α-GMP6A (OriGene AM26432AF-N), 1:500
- Rabbit α-Syp (LSBio aa225-253), 1:1000
- Mouse α-Snap25 (Abcam ab66066), 1:1000
- Mouse α-Gria2 (Abcam ab106515), 1:500
- Mouse α-Cntn1 (Abcam ab105582), 1:500

- Mouse α-Syt1 (Abcam ab77314), 1:2000

Secondary antibodies

- Donkey α-mouse IgG AF568 (ThermoFisher A10037), 1:1000
- Donkey α-rabbit IgG AF568 (ThermoFisher A10042), 1:1000
- Donkey α-rabbit IgG AF488 (Abcam ab150073), 1:1000
- Goat α-Rat IgG AF568 (ThermoFisher A21434), 1:1000

Cells in monolayers were washed once with PBS, then fixed for 15' at RT in 4% PFA. They were then permeabilised for 1h at RT in permeabilising solution, then blocked for at least 3h at RT in blocking solution. Primary antibodies were diluted in blocking solution and incubated overnight at 4ºC. Cells were then washed 3 times with PBST, then incubated for 45' with secondary antibodies diluted in blocking solution, at RT in the dark. DAPI (ThermoFisher D3571) to visualise cell nuclei was added together with the secondary antibodies at a final concentration of 1µg/ml. Cells were washed again 3 times with PBST and then imaged in PBS. Between imaging, plates were kept at 4ºC to maintain fluorescence, up to 10 days.

2.5.2 Imaging with Celigo Cytometer

Whole well images were obtained using the Celigo Cytometer (Nexelcom). Table 2.7 below summarises the laser conditions of the machine

Channel	**Excitation**	**Dichroic**	**Emission**
Blue	377/50	409	470/22
Green	483/32	506	536/40
Red	531/40	593	629/53

Table 2.7 – Celigo Cytometer laser settings.

2.6 Library selection, cloning, preparation for NGS and analysis

2.6.1 Gene Ontology analysis

Gene Ontology analysis of the genes included in the library was performed using the online tool on the PANTHER DB website using standard settings (http://pantherdb.org) (Mi *et al.*, 2019). Analysis was performed for both biological processes and molecular function. Terms were ranked on -log(p value). P-value is calculated by PANTHER as the probability of a GO term being represented in a given list considering its frequency of representation in a background genome of reference.

2.6.2 Oligo library cloning and preparation for NGS

Oligoes were ordered as a ssDNA pool from TWIST Bioscience and resuspended to a final concentration of 1ng/μl in AE buffer (Qiagen) (10 mM Tris-Cl, 0.5 mM EDTA; pH 9.0). PCR was used to amplify and generate dsDNA oligoes. PCR mix and conditions are represented in tables 2.8 and 2.9. The final amplicon structure is a 79-mer oligo with the sequence:

5'-GCAGATGGCTCTTTGTCCTAGACATCGAAGACAACACCG-N_{19}
GTTTTACAGTCTTCTCGTCGC-3'

where N_{19} represents each specific gRNA. Primers were taken from Dr. Kosuke Yusa's protocol (Koike-Yusa *et al.*, 2014).

12.5 ul	Q5 Hot Start High-Fidelity Master Mix (2X)
1 ul	Library oligoes (1ng/ul)
1.25 ul	F-mer U1 + L1 Primer (10uM)
10.25	dH20

25ul total	

Table 2.8 – Oligo library 1st PCR mix

98C	10s	
98C	10s	X10
64	15s	
72	15s	
72	2min	

Table 2.9 – Oligo library 1st PCR conditions

Cycle number was determined by qPCR saturation curves from 5µl of reaction supplemented with 1X SYBR and run in the Light Cycler 480.

The PCR product was cleaned with Oligonucleotide Removal Kit (Qiagen). Correct band size and purity was confirmed by gel electrophoresis through a 2% agarose gel. Fragments and pKLV2(W-)U6sgRNA5-sEF1aBFP-W library backbone were digested overnight with Bbs1 (NEB) at 37ºC, conditions are reported in the tables 2.10 and 2.11 below.

5ug	pLV backbone
5ul	NEB buff 2
0.5ul	BSA 100X
5ul	Bbs1
Up to 50ul	dH20

Table 2.10 – Library backbone digestion

40ul	Annealed and cleaned-up library
10ul	NEB buff 2 (10X)
1ul	BSA 100X
5ul	Bbs1 enzyme

Up to 100ul	dH20

Table 2.11 – Library dsDNA oligoes digestion

Linearized backbone was run through a 1% agarose gel and purified using Zymoclean Gel DNA recovery Kit (Zymoclean, D4001). Digested dsDNA oligoes were run through a 20% PAGE gel to isolate the middle fragment of 26bp and containing the gRNA sequence to be ligated from the flanking regions necessary for amplification only. The 26bp fragment was cut out and eluted from PAGE gel by incubation in PAGE elution buffer (10mM magnesium acetate tetrahydrate, 0.5M ammonium acetate, 1mM EDTA pH 8.0) for 3h at 37ºC in a 800rpm shaker.

Eluted solution was purified with QIAquick nucleotide removal kit (Qiagen) but using MiniElute columns (Qiagen).

Ligation was set-up overnight at 16ºC using 100ng of backbone and 1:4 molar ratio of backbone to insert, using fresh T4 Ligase buffer and T4 Ligase (2M units/ml). Ligated product was eluted in water using ZymoResearch DNA clean and concentrator kit. 2ng were electroporated into NEB electrocompetent bacteria 10-β. The pool of electroporated bacteria was inoculated into 500ml LB broth supplemented with Ampicillin (100 µg/mL) and incubated overnight at 37ºC. Library plasmid pool was isolated from bacteria through Maxiprep (Qiagen).

2.6.3 Plasmid and screen library preparation for NGS

Plasmid library samples and screen library samples from dsRED+ve and dsRED-ve cells shared the same amplicon and thus were prepared following the same protocol. 1ng plasmid DNA was used for plasmid technical replicates. 181.1ng gDNA were used for

cell sorted samples to achieve a gRNA coverage of 100X. The amplicon structure is shown below:

5'-TCTTGTGGAAAGGACGAAACACCGN$_{19}$TTAAGAGC----146bp----CAGAGGACGACGCTCTAGGGATTTGT-3'

Where N_{19} represents the gRNA sequences, and the stretches in red are the regions of homology for primer annealing for the 1st round of PCR. This round is needed to add sequence adaptors that will anneal to Illumina Nextera primers.

PCR1 primers are shown below. In red, overlapping regions to the amplicon above.

F: 5'-TCGTCGGCAGCGTCTCTTGTGGAAAGGACGAAACACC-3'

R: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAAATCCCTAGAGCGTCGT-3'

PCR was run with Q5 Hot Start High Fidelity Polymerase 2x mix and primers at 0.2µM in a final volume of 50µl with the conditions shown in Table 2.12. Cycle number was determined by qPCR saturation curves from 5µl of reaction supplemented with 1X SYBR and run in the Light Cycler 480.

98C	30s	
98C	10s	X13
64	15s	
72	20s	
72	2min	

Table 2.12 – NGS library preparation, 1st PCR conditions.

PCR product was gel-purified from a 2% agarose gel with Zymoclean Gel DNA recovery Kit. 1ng was then used as template for the 2nd PCR reaction, required to add Illumina Nextera adaptors for sequencing. The new template structure is shown below:

5'-TCGTCGGCAGCGTCTCTTGTGGAAAGGACGAAACACCGN19TTAAGAGC----146bp----
CAGAGGACGACGCTCTAGGGATTTGTCTGTCTCTTATACACATCTCCGAGCCCACGAGAC-3'

Illumina Nextera primers from Nextera® XT Index Kit anneal to the blue and orange stretches. i5 and i7 indexes included in the adapters to allow multiplexing are incorporated in the primer pairs. The final amplicon including indexing structure is shown below:

5'-
AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTCTCTTGTGGAAAGGACGAAAC
ACCGN19TTAAGAGC----146bp----
CAGAGGACGACGCTCTAGGGATTTGTCTGTCTCTTATACACATCTCCGAGCCCACGAGAC[i7]ATCTC
GTATGCCGTCTTCTGCTTG-3'

Where Nextera primer sequences are underlined.
PCR was performed with 2X Kapa HiFi HotStart ReadyMix (KAPA KK2602) and 10X Nextera Primers to a final volume of 50µl. The 2nd PCR conditions are outlines in tables 2.13 below. Cycle number was determined by qPCR saturation curves from 5µl of reaction supplemented with 1X SYBR and run in the Light Cycler 480.

98C	30s	
98C	10s	X7
64	15s	
72	20s	
72	2min	

Table 2.13 – NGS library preparation, 2nd PCR conditions.

PCR products were purified with Agencourt AMPure XP (SPRI) magnetic beads (Beckman Coulter, A63881. Individual libraries quality and quantification was

performed with Qubit and TapeStation. They were then pooled to equimolar amounts to a final concentration of 10μM and sent for NextSeq.

2.6.4 NextSeq

NextSeq was performed at the EMBL Genomics Core Facility by Dr. Vladimir Benes's team and with the help of Dr. Jelena Pistolic and Dr. Dinko Pavlinic. The run was performed as a single end read, paired index deconvolution with custom read primer and custom i5 read primers reported below, supplied at a 100μM concentration.

Custom read primer: 5'-TCGTCGGCAGCGTCTCTTGTGGAAAGGACGAAACACCG-3'

Custom i5 read primer: 5'-CGGTGTTTCGTCCTTTCCACAAGAGACGCTGCCGACGA-3'

Read run was interrupted after with 19 cycles to selectively read the unique gRNA sequence within the amplicon.

2.6.5 MAGeCK analysis of NextSeq results

Analysis of gRNA read counts from .fastaq files was performed by Dr. James Ashmore using the default MAGeCK settings (Li *et al.*, 2014). gRNA read counts within each sample were first normalised against total read counts per sample. Then, a negative binomial model was used to assign a p-value to each gRNA based on the difference in read counts between treatment sample and control sample, and the p-value was used to rank each gRNA. An α-RRA algorithm was then used to associate the behaviour of gRNAs targeting the same gene and identify, by identifying skews in gRNAs from the uniform null distribution. The significance of this skew was used to rank enriched and depleted genes. The treatment vs control sample comparison was performed between iN (treatment) and plasmid (control), iN (treatment) and non-neuronal cells (control), and non-neuronal cells (treatment) and plasmid (control).

2.7 Gene Expression analysis

RNA was isolated from whole wells or sorted cells RNAeasy Mini Kit (Qiagen 74104) following manufacturer instructions. After quantification, it was converted into cDNA using SuperScript™ VILO™ Master Mix (Invitrogen 11755050), following manufacturer instructions. 2ng of cDNA (estimated from RNA quantification) were used per qPCR reaction. qPCR reactions were prepared using Brilliant III Ultra-Fast SYBR PCR 2X mix (Agilent) in final volume of 10μl within 384-well plates. Primers were added to a final concentration of 0.25μM. Plates were run in the LightCycler480. Table 2.14 below lists the primers used for gene expression qPCR in this study.

Primer name	**Sequence**
Gria2 F	GGGGACAAGGCGTGGAAATA
Gria2 R	GTACCCAATCTTCCGGGGTC
Map2 F	CAGAGAAACAGCAGAGGAGGT
Map2 R	TTTGTTCTGAGGCTGGCGAT
Snap25 F	TTCATCCGCAGGGTAACAAA
Snap25 R	GTTGCACGTTGGTTGGCTT
Stmn3 F	AGCACCGTATCTGCCTACAAG
Stmn3 R	TGGTAGATGGTGTTCGGGTG
Tubb3 F	CAGATAGGGGCCAAGTTCTGG
Tubb3 R	GTTGTCGGGCCTGAATAGGT
Nestin F	CTCAGATCCTGGAAGGTGGG
Nestin R	GCAGAGTCCTGTATGTAGCCA
Fibronectin F	CGAAGAGCCCTTACAGTTCCA
Fibronectin R	ATCTGTAGGCTGGTTCAGGC
TBP F	GGGGAGCTGTGATGTGAAGT
TBP R	CCAGGAAATAATTCTGGCTCA

Table 2.14 – gene expression qPCR primers

2.8 Molecular Biology

2.8.1 sgRNA cloning

Individual sgRNA cloning made use of the genome-wide sgRNA library generated by Tzelepis and colleagues (Tzelepis *et al.*, 2016) containing 19bp sequences. These sequences were ordered as individual complementary oligoes, with 5' adaptors to allow ligation into the pLV-U6-gRNA backbone as follows:

+ Strand: $CACCGN_{19}$ – where N_{19} is the stretch on the +ve strand from the library itself
- Strand: $AAAC(N_{19}(rc))C$ – where $(N_{19}(rc))$ is the reverse complement of N_{19}

Oligoes were mixed to equimolar amounts, heated up to 95ºC and allowed to cool gradually to create a double stranded fragment with ligation-ready overhangs.
The lentiviral backbone was digested with BbsI (NEB R0539) to linearize it, then gel-purified and ligated with the annealed gRNA fragment at a 3:1 oligo : vector ratio. Ligation products were transformed into chemically competent bacteria. Individual colonies were inoculated and amplified to isolate plasmids via miniprep. Plasmid integrity was confirmed using restriction digestion patterns. Presence of the correct sgRNA was confirmed via Sanger Sequencing.

2.8.2 Stxbp1 cDNA cloning for overexpression

Adult mouse whole-brain total RNA was obtained by processing a $0.5cm^3$ section of 2-month old mouse brain using the RNAeasy Mini Kit (Qiagen 74104) following manufacturer instructions. RNA was then converted into cDNA using SuperScript™ VILO™ Master Mix (Invitrogen 11755050), following manufacturer instructions. 50ng of cDNA (estimated from RNA quantification) were used per PCR reaction as template for the amplification of Stxbp1a/b cDNAs. PCR primers and conditions are reported below in tables 2.15 and 2.16. The primers contain EcoRI restriction sites for digestion and ligation into the pLV-tetO lentiviral backbone.

Stxbp1_F	TAATAAGAATTCTACCACCATGGCCCCCATTGGCCTCAA
Stxbp1_R_IsoformA	TTATTAGAATTCTCACTCCATTGTTGGAGCCTGA
Stxbp1_R_IsoformB	TTATTAGAATTCGCAGCTTATTTTTTAACTGCTTATTTCTTCATCTG

Table 2.15 – Stxbp1 cDNA amplification primers

95C	3'	
95C	10s	x20
63C	10s	
72C	30s	
72C	5'	
4C	cooling	

Table 2.16 – Stxbp1 cDNA amplification PCR conditions

PCR products were run through a 1% agarose gel using electrophoresis and confirmed the amplification of fragments of the correct expected size (1.6kb). After gel purification, the fragments were digested for 1h at 37ºC with EcoRI to create compatible ends for ligation. At the same time, the pLV-tetO-Ascl1 plasmid was digested with EcoRI to excise Ascl1 cDNA and gel-purified to obtain a digested backbone ready for ligation. Inserts and backbone were ligated to a molar ratio of 3:1 overnight at 16ºC, and then transformed into competent bacteria. Colonies were selected and plasmids isolated via mini-prep. Correctly cloned vectors were verified by restriction enzyme digestion. cDNA correct sequence was verified with Sanger sequencing.

2.8.3 TIDE

PCR amplicons for TIDE analysis were designed using Primer Blast software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) so that they would amplify a genomic region of about 400bp containing the gRNA sequence roughly equidistant

from both ends. They were selected for specificity and similar annealing temperatures, and verified to amplify a single genomic product by gel electrophoresis. PCR conditions are shown in Table 2. below. PrimeStar 2X MasterMix (Takara Clontech) was used to amplify the fragments, and 50ng genomic DNA were used as template.

95C	3'	
95C	10s	x20
Anneal T	10s	
72C	30s	
72C	5'	
4C	cooling	

Table 2.17– TIDE PCR conditions

Amplicons were column-purified using Zymoclean DNA Clean and Concentrator Kit (Zymoclean, D4003) and sent for Sanger Sequencing at Source Bioscience, using the forward amplification primer as a sequencing primer. Sequencing mixes contained 1 ng/µl per 100 bp size of PCR product, primer was provided separately at 1.6 pM. Chromatograms from edited and wild-type cells were uploaded on the TIDE website (https://tide-calculator.nki.nl) with default settings (Brinkman et al., 2014).

<u>Primer pairs for Ascl1 gRNA #1 TIDE:</u>

TIDE_Ascl1_g1_F TCGTCCTCTCCGGAACTGAT

TIDE_Asck1_g1_R TCCTCTGGGCTAAGAGGGTC

<u>Primer pairs for Stxbp1 gRNA #3 TIDE:</u>

TIDE_Stxbp1_g3_F AGGTGTTATCCTTAGGAACCAGAG

TIDE_Stxbp1_g3_R TGAGCTCACTCCCCAGGTTG

2.9 General molecular biology techniques

2.9.1 Ligation

Ligation was performed at 16ºC for 1h (gRNA ligation) or overnight (PCR product ligation). 50ng of linearized purified vector were mixed with the insert at the specified ratio. 1µl T4 Ligase (NEB) and 1µl 10X T4 Ligase buffer (NEB), and water was supplemented to a final volume of 10µl. 5µl of ligation product were used for transformation.

2.9.2 Transformation

Transformation was performed in chemically competent homemade DH5α *E.coli.* Bacterial aliquots of 25µl were thawed on ice for a couple of minutes before being supplemented with DNA. 5µl of ligation product was used to transform ligations, 1ng of plasmid DNA was used to transformed closed plasmids that needed to be amplified. After 2 minutes of incubation on ice, bacteria were heat-shocked at 42ºC for 45 seconds, then returned on ice for 3 more minutes. 4 volumes of LB broth were then added to the samples (100µl for the 25µl aliquots) and tubes were incubated for 1h at 37ºC to allow expression of the antibiotic resistance gene carried within the plasmid. The total volume was then plated on LB agar plates containing the appropriate antibiotic for selection. Plates were incubated overnight at 37ºC before colony picking.

2.9.3 Plasmid amplification and verification

A single colony was picked from bacteria plates and inoculated within LB broth containing the appropriate antibiotic. Colonies were inoculated in 3ml for mini-prep if the plasmid was being constructed, or in 50ml for midi-prep for plasmid amplification. They were incubated overnight at 37ºC within shakers (200rpm) to

allow bacterial expansion. The next day, they were processed with Qiagen Miniprep or Midiprep kits following manufacturer instructions to isolate plasmid DNA.

Plasmid integrity was verified with restriction enzyme digestion followed by gel electrophoresis. 100ng of plasmid DNA was digested for 1h at 37ºC in a final volume of 10µl with appropriate restriction enzymes and buffers (NEB), then run on 1% agarose gel to examine band pattern.

For gRNA cloning and PCR product cloning, Sanger sequencing from Source Bioscience was used to verify correct gRNA/product sequence. Sequencing mixes contained 100ng/of plasmid DNA. Primer was provided separately at 1.6 pM.

2.10 List of plasmids

Name	**Source**
pLV_FUW-tetO-*Ascl1*	Addgene plasmid # 27150
pLV_FUW-tetO-*Brn2*	Addgene plasmid # 27151
pLV_FUW-tetO-*Myt1l*	Addgene plasmid # 27152
pLV_FUW-M2rtTA	Addgene plasmid # 20342
pLV_*hSyn1-dsRed*	Addgene plasmid # 22909
psPAX2	Addgene plasmid #12260
pMD2.G	Addgene plasmid #12259
pKLV2(W-)U6sgRNA5-sEF1aBFP-W	Dr. Kosuke Yusa

Table 2.18 – plasmids used in this study

2.11 List of gRNAs

2.11.1 Control gRNAs

Name	**gRNA sequence**
Zeomycin (non-targeting gRNA)	GCGAACTTCCGGGACGCCTC
*Pecam1*_intron1	CCTCTTGGGTTCTTTGTGAC
*pg53*_g2	AATAAGCTATTCTGCCAGC

Table 2.19 – Control gRNAs

2.11.2 gRNAs in Chapter 4

Name	**gRNA sequence**
B3gnt1_g1	TACGGATCGCGGAGACGGC
B9d1_g2	CAAGATGTACGGCAAGCAC
Blvrb_g2	GAAGCTGTCATCGTGCTAC
Camk4_g3	ACTTACAGCAACGGCCTCC
Ddx41_g4	CGAAGACGCAAGGGCGCTA
Fez1_g1	ACATCTCCTCATCACCTGG
Nid1_g5	GTGTGCCCACGATGACGCC
Rcn3_g2	CCTACGACACGGACCGCGA
Casc3_g3	CCGAAGAATCCGGAAACCT
Casc3_g4	CCAGGAATTACTCTCGATC
Ccng2_g1	GATACCAACCTCGGGAAAA
Ccng2_g4	GTGACAAAATACAATCGCG
Cnrip1_g4	AGACAACTCTCTCCCCGTC
Cnrip1_g5	GTGTAGCCCCAACCAAAAG
Gtf2i_g2	GAGGAATGCCGTACGTCGA
Gtf2i_g3	GAGGGTACGGGATCGTCAC
Ier3_g1	CTTCGACCCTCTCCCGGAG

Ier3_g2	GAAGTGTTCAAACGCGCGG
Os9_g5	TGGGACGACGAAACAGCCA
Os9_g2	ACTGACCATTCGCACGTCA
Sf3a1_g1	GGTAAATAATGCCGACCAC
Sf3a1_g2	CGACGATGTTCCTGACCTC
Sf3a1_g3	CTTGTGGCGGTAGTAGGCA

Table 2.20 – gRNAs used in Chapter 4

2.11.3 gRNAs in Chapter 5

Name	**gRNA sequence**
Kmt2b_g1	ACGCGTAGACGGCGCCACT
Kmt2b_g2	TGCGACGTCTTGGAGCAGG
Kmt2b_g3	CCGACGCTGCGTCAAACAG
Dll1_g1	ACGAGCACTACTACGGAGA
Dll1_g2	AAACCGGTAAGAGTACCGG
Dll1_g3	CTCCCGATGACCTCGCAAC
Hes6_g1	CTACCCTCACGCAGCGGCA
Hes6_g2	CAAGCGAGCGCTTCGCTGC
Hes6_g3	GCGCCCTGCACGCGCCTCA
Lmo2_g1	GCAGCCACCACATGTCAGC
Lmo2_g2	GCGGTCCCCTATGTTCTGC
Lmo2_g3	GAAAGCCATCGACCAGTAC
Mfng_g1	TTCTGCATCAACCGCCAAC
Mfng_g2	GACAACTATGTGAACCCCA
Mfng_g3	AGAACAGTTGGTGACCACG
NeuroD4_g1	AAGATAGAGACTCTTCGAC
NeuroD4_g2	ACTCTCCTAAGATTATCCA
NeuroD4_g3	CAATGCTAGAGAACGGACC

Olig2_g1	GTAGCTCGGAGCTCAGCTC
Olig2_g2	CGCGTCAGGCGCGCATCCC
Olig2_g3	CCGGGGACAAACTGGGCGG
Pou3f4_g1	TGACCAGTCTTAGCGACGG
Pou3f4_g2	GACGCGAAGATCTGCAACT
Pou3f4_g3	ACTAACCATCCGAACGCCT
Rfx1_g1	GATGTGAACTTACCCGATA
Rfx1_g2	GATAGGCGTCTCCGGGTAC
Rfx1_g3	AGTAGCCGGCGTGCTGACC
Tcfl5_g1	TTGAGCGCCACGCCCGACT
Tcfl5_g2	CCGACCGTACGCGTCCGTC
Tcfl5_g3	CGGGCCTCCGGTGCGCCGT
Zfp238_g1	AGCTCGGTGTGCCCGGAAC
Zfp238_g2	GCTTGAATTCATGTACGAA
Zfp238_g3	GCTGCTAGCACGTCCTCAA
Zic1_g1	GCCGTACTGTTCCGGCCGT
Zic1_g2	AACCACGTTGGGCGAAGCG
Zic1_g3	ACGGCCATACGGACGCCGC
2700081O15Rik_g1	GGTAGGGGGGTACTCCCCC
2700081O15Rik_g2	CCAGCGGCGAGGGGTCCGA
2700081O15Rik_g3	CTCAGTGGCACCTCGGCGC
Anxa3_g1	CTATCCGGAAAGCGATCAG
Anxa3_g2	CAGTTGCTTCGCATCAAAC
Anxa3_g3	CCTAACAACCAGGTCAAGC
Cnn2_g1	GAACTCCGAAGCTGGATAG
Cnn2_g2	CAACCGCTCTATGCAGAAC
Cnn2_g3	AGCTACGGCATGAACCCTG
Dcaf5_g1	TCAATTTCACGACGCACCA
Dcaf5_g2	TGACTGGTTAGCGTAGTCG

Dcaf5_g3	CATCGGTCCATCGTGAACC
Fam171b_g1	AAACACCTCCACAACGGCT
Fam171b_g2	CGTCTTGACATCTCTGCCG
Fam171b_g3	TGTTTGGGTGCATCACGGT
Gja1_g1	TGACTCAACCGCTGTCCCC
Gja1_g2	TGTTGAGTGTTACAGCGAA
Gja1_g3	TAGCAGACATTTTCACAAC
Stxbp1_g1	AGCCGACGATCCAACAATG
Stxbp1_g2	AATCCGATACTGGAACGCC
Stxbp1_g3	GTGTCGCAGCGCAATCCAC
Vmac_g1	CCATGTTGTTGCCCAGGGG
Vmac_g2	ATGGGCCTGGCCCACCCCT
Vmac_g3	GGCCCATCAGGAAGTAGCT

Table 2.21 – gRNAs used in Chapter 5

Chapter 3 Optimising the Induced Neuron conversion system for a CRISPR/*Cas9* KO screen

3.1 Introduction

In order to perform a CRISPR/*Cas9* KO screen in the iN system, I needed to address two major questions: how to isolate the iNs from the culture system, and how to select a subset of candidate genes to screen for.

The pan-neuronal marker *Tubb3* and the post-mitotic neuronal marker *Map2* are two of the most common markers used to identify induced neurons in the conversion

contexts, both in mouse and human cells, as two of the earliest markers to be upregulated and marking virtually all neurons, irrespective of their subtype (Vierbuchen *et al.*, 2010; Pfisterer *et al.*, 2011a; Pang *et al.*, 2011; Son *et al.*, 2011).

These markers have been used to confidently quantify iNs from immunofluorescence experiments, as they highlight the full morphology of the neuronal cells and can be used to assess cell body size but also neurite lengths – a criterion used in the original 2010 paper by Vierbuchen and colleagues to define a neuronal cell.

The proteins though are intracellular and their application as markers for flow cytometry can be quite tricky (Turaç *et al.*, 2013). Intracellular flow cytometry in general requires cells permeabilization and is prone to issues with antibody non-specificity (Demaret *et al.*, 2017), and would not have been very straightforward to adapt to such a heterogeneous system.

The Wernig lab has made use of a *Tau-eGFP* reporter line, generated first to select ESC-derived post-mitotic neurons (Wernig *et al.*, 2002). *Tau-eGFP* was confirmed to selectively label *Tubb3*+ve/*Map2*+ve induced neurons and hence used to isolate iN for flow cytometry for their RNAseq experiments in subsequent papers (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013; Treutlein *et al.*, 2016).

We were unable to obtain the *Tau-eGFP* mESC line (Tucker, Meyer & Barde, 2001) that would have allowed us to obtain chimaeric MEFs through morula aggregation.

I set out to identify alternative reporters, testing a subset of genes specifically expressed in iNs for which I could find flow-cytometry appropriate antibodies. I then moved on to characterise the activity of the human *Synapsin1* promoter as a specific marker for iNs.

This promoter is broadly active in neuronal cells of the central and peripheral nervous system, but not in other neural cells, and was described to be selectively active in various mammalian neuronal cell lines (Thiel, Greengard & Sudhoft, 1991). A 470bp fragment of this promoter was shown to selectively mark neurons in rats (Kügler, Kilic & Bähr, 2003) and mice (McLean *et al.*, 2014). Moreover, it was also used to mark human ESC-derived neurons *in vitro* (Patzke *et al.*, 2015). In the MEF to iN conversion system, it had been previously used to mark the most mature (morphologically and

functionally) iNs in the heterogeneous mix of reprogrammed cells, (Black *et al.*, 2016; Adler *et al.*, 2012) to allow identification of neurons for electrophysiological studies. Nonetheless, its temporal activation and its use for flow-cytometry isolation of iNs has not been addressed in those two studies. It presented as a good candidate marker for its ease of use, and I decided to characterise its expression in this system.

To perform the genetic screening, I decided to take advantage of the CRISPR/*Cas9* system established in the lab. The recent advances in this technology made it highly suitable for genetic screenings, taking over the more traditional RNAi approach that was burdened with variable knockdown (KD) of transcripts and high off-target effects (Jackson & Linsley, 2010; Boutros & Ahringer, 2008).

Genome-wide CRISPR/*Cas9* screens have been successfully performed in murine and human cells (Zhou *et al.*, 2014b; Shalem *et al.*, 2014; Wang *et al.*, 2014; Koike-Yusa *et al.*, 2014) and allowed the identification of specific gene networks and pathways. Further improvements, such as optimisation of the gRNA scaffold, lentiviral gRNA delivery and controlled and homogeneous *Cas9* expression have enhanced the power of the technology and allowed the identification of broader gene subsets with statistical robustness (Tzelepis *et al.*, 2016).

Considering the low neuronal yields of the conversion and the processing times required to harvest and sort the iNs from a large-scale culture, I decided to select a library of candidate genes that are likely to play a role in the conversion, instead of using a genome-wide gRNA approach. Using gene expression as a predictor of gene function, I made use of sc-RNAseq dataset from the Wernig lab (Treutlein *et al.*, 2016) to select a subset of genes to screen against. The gRNA sequences per each gene were selected from the mouse genome-wide library previously used in the lab for a different screening project (Kaemena, Beniazza et al., submitted) obtained from our collaborator Dr. Yusa (Tzelepis *et al.*, 2016)

3.1.1 Aims of the chapter

The aims of this chapter are to establish the iN system as a platform for performing the CRISPR/*Cas9* screen. I started by verifying the compatibility of the set-up with the *Cas9* expressing MEFs and *Cas9* activity.

I then tested a series of candidate genes as selective reporters for isolating neurons from the heterogeneous culture. I selected the human *Synapsin1* promoter and validated that its activity is specific for iN and shows very little leakiness from MEFs in conversion medium.

The second part of the chapter is focused on generating a gRNA pooled plasmid library to perform the screen. Based on neuronal yields and recommended coverage from previous works from our collaborator Dr. Kosuke Yusa and previously published CRISPR/*Cas9* screens, a targeted pool of 1203 genes was selected based on their expression across the conversion. gRNA sequences were cloned as a pool into the appropriate vector for expression, and the quality of this library was assessed with NGS.

3.1.2 Notes

1) The gRNAs used in this chapter target the Zeomycin resistance gene, *Pecam1* (intron 1) and *p53*.

   The Zeomycin gRNA is used as a control for the *Cas9* genome-scanning activity that has the potential to impact cell behaviour, and it is referred to as a non-targeting control, since the mouse genome does not contain a Zeomycin resistance gene. The *Pecam1* intron 1 gRNA – referred to as *Pecam1* or Pe1 – targets the 1st intron of the *Pecam1* gene. This control, targeting an intronic region, is made to control for *Cas9* activity and the effects of DNA double strand break (DBS) on the conversion and I include it when quantifying the neuronal conversion efficiency by flow cytometry. Though *Cas9* activity and DBS will always be specific for each gRNA's genomic location, the *Pecam1* intron1 gRNA has been previously used in the lab as a neutral control in MEF

to induced pluripotent stem cells (iPSCs) reprogramming experiments, as it was shown to not impact the reprogramming efficiency. I have verified that it acts as a neutral control for MEF to iN conversion experiments as well, by comparing the neuronal conversion efficiencies in non-transduced samples and samples transduced with either Zeomycin gRNA or *Pecam1* gRNA.

P53 KO is known to induce MEF proliferation. The *p53* gRNA used has been verified to be functional previously in the lab and has been shown to induce the proliferation phenotype expected. I chose to test *p53* KO to verify the robustness of the *Cas9*-iN system, because this KO has a clear and observable phenotype that I could easily monitor and confirm visually. I also wanted to verify the role of *p53* KO based on what has been previously reported about *p53* in neuronal conversions – explained in more detail later on.

Below are the sequences for all three gRNAs:

a. Zeomycin: GCGAACTTCCGGGACGCCTC
b. *Pecam1*: CCTCTTGGGTTCTTTGTGAC
c. *P53* gRNA #2: AATAAGCTATTCTGCCAGC

3.2 Results

3.2.1 Setting up the iN conversion system with our MEF line

The first thing I wanted to verify is whether our *Cas9* expressing MEFs were capable of conversion with efficiencies and neuronal qualities similar to the original reported system by Vierbuchen and colleagues in 2010.

As shown by the schematic in Fig. 3.1a, the lentiviral vectors, obtained through Addgene, allow the expression the 3 Wernig transcription factors (TFs) under the control of the tetOn promoter, thus requiring the addition of *rtTA* and doxycycline to activate TF expression. *rtTA* was expressed separately by co-transducing the cells with the pLV-hUb-*rtTA* vector, leading to constitutive *rtTA* expression thanks to the human *Ubiquitin* promoter.

I transduced the MEFs with the 3 TFs and *rtTA* at MOI5 and induced their expression with doxycycline 24h after transduction (Fig.3.1b). I then tracked the appearance of neuronal cells by the expression of markers *Tubb3* and *Map2* at day 5, 15 and 22 of the conversion by immunofluorescence. As previously reported, I observed cells with TUBB3/MAP2 expression with immature neuronal morphology, thin projections and little branching as early as day 5. By day 15, the somas of the neuronal cells have started to cluster together, and the clusters are interconnected by a network of highly branched projections. At day 22, the clusters have grown bigger and the connections thicker (Fig.3.1.c)

I quantified the neuronal purity in the culture comparing the numbers of neurons vs the overall nuclei present via immunofluorescences. Although the stochastic nature of lentiviral transduction and conversion both make the system quite variable, I obtained an average of 18% of cells of neuronal identity – as assessed by the expression of TUBB3. The efficiency of conversion was estimated based on the initial number of seeded MEFs. This resulted in a neuronal yield of 23% on average (Fig.3.1d-e).

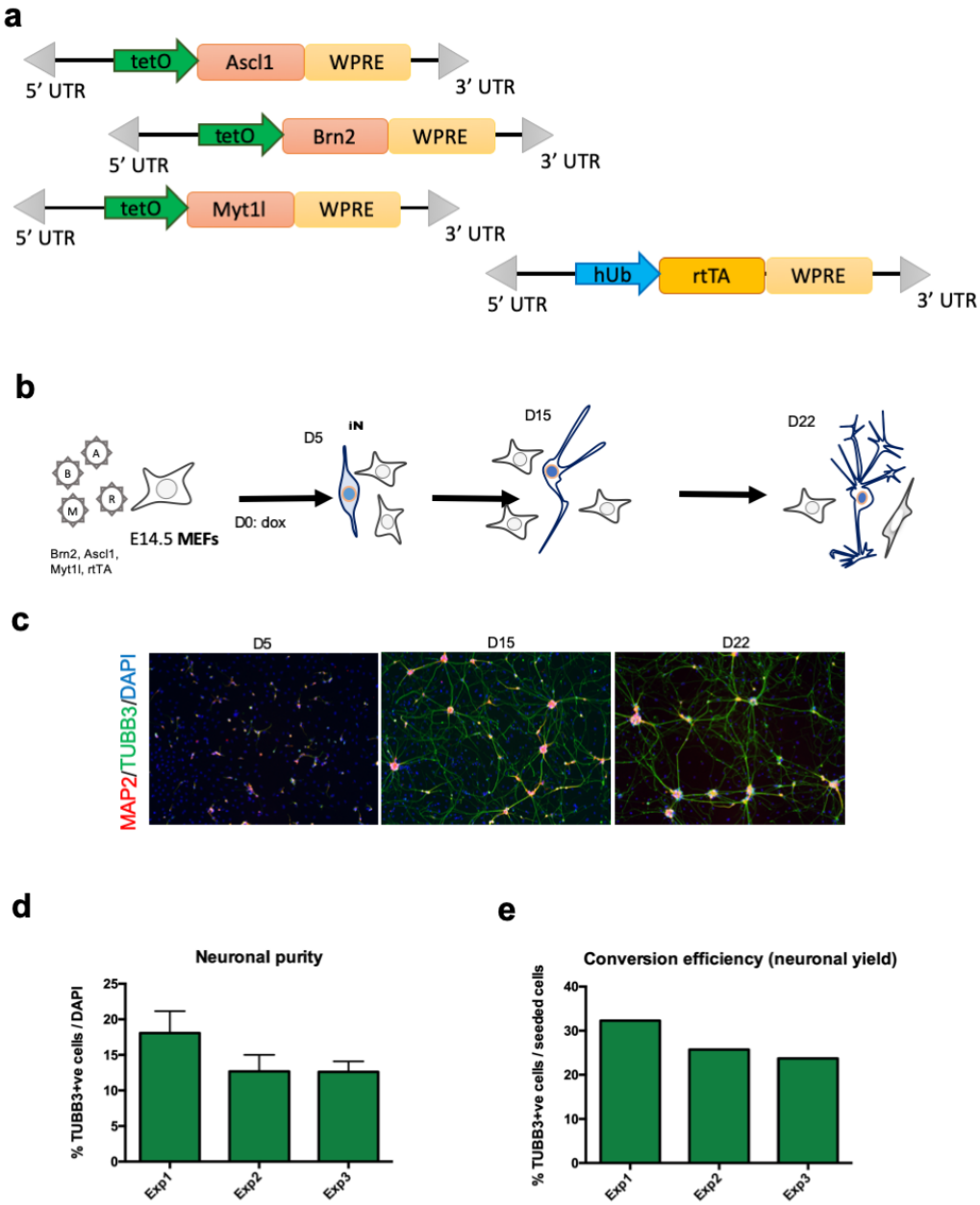


Fig. 3.1 – the MEF to induced neuron conversion system. (a) Schematic representation of the lentiviral expression vectors for transduction of the converting TFs and *rtTA*. Promoter sequence tetO, hUb; cDNAs: *Ascl1*, *Brn2*, *Myt1l*, *rtTA*; WPRE: Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. **(b)** Explanatory diagram of a conversion experiment. E14.5 MEFs are transduced with the lentiviral particles one day after seeding. At day 0, doxycycline is added to the culture medium to induce expression of *Ascl1*, *Brn2* and *Myt1l*. Converting cells progressively acquire neuronal morphology and expression of neuronal markers. **(c)** Immunofluorescence at day 5, 15 and 22 post-induction of the TFs shows cells with neuronal morphology that progressively becomes for complex, and express

the pan-neuronal marker TUBB3 and the post-mitotic neuronal marker MAP2 **(d)** Quantification of the neuronal purity at day 15 of the conversion, shown as % of neuronal cells over total number of cells per field. 10 fields were analysed per experiment, error bars represent STDEV. **(e)** Quantification of the conversion efficiency towards neurons at day 15 of the conversion. The conversion efficiency is calculated as the total number of neurons per well over the initial number of seeded MEFs and represents the neuronal yield of the conversion. Total number of neurons are estimated based on 10 fields per experiment. The yields of three individual experiments are shown separately to highlight variability.

To verify the compatibility of this system with our platform for CRISPR-*Cas9* KO, I decided to knock-out *p53* and monitor the effects on the conversion efficiency and neuronal yield. *P53* was chosen based on the described effect of its KO on MEF proliferation, and on availability of tested gRNAs. In the context of iNs, *p53* knock-down has been shown to increase the efficiency of conversion from human fibroblasts to induced dopaminergic neurons (Jiang *et al.*, 2015), and its knock-out has been shown to be sufficient for the conversion of human and mouse fibroblasts into induced neurons (Zhou *et al.*, 2014a).

I transduced *Cas9*-MEFs with the gRNA against *p53,* together with the TFs at day -1 (as control, I used the non-targeting gRNA against Zeomycin). This resulted in an initial wave of MEF over-proliferation compared to the control gRNA, as expected. Doxycycline was added at day 0, and by 48h post-induction, the proliferation, assessed visually, halted. At day 15 cells with neuronal morphology and expression of neuronal markers were present in both samples, confirming that genome editing by CRISPR/*Cas9* KO does not impact the ability of cells to convert into neurons. The purity of neurons in the culture was very mildly decreased upon *p53* KO (Fig.3.2c), while the slight increase in yield is reasonably due to the initial cell proliferation. This is consistent with previous reports in human cells showing that the neuronal yield of the conversion can be increased by activating TF expression after a few days of transduction, to allow cell proliferation and increase the number of cells harbouring the TFs expression cassettes and thus able to reprogram (Pereira et al.,2014). I did

not observe an increase in neuronal conversion efficiency nor total neuronal yield in this system and under these conditions.

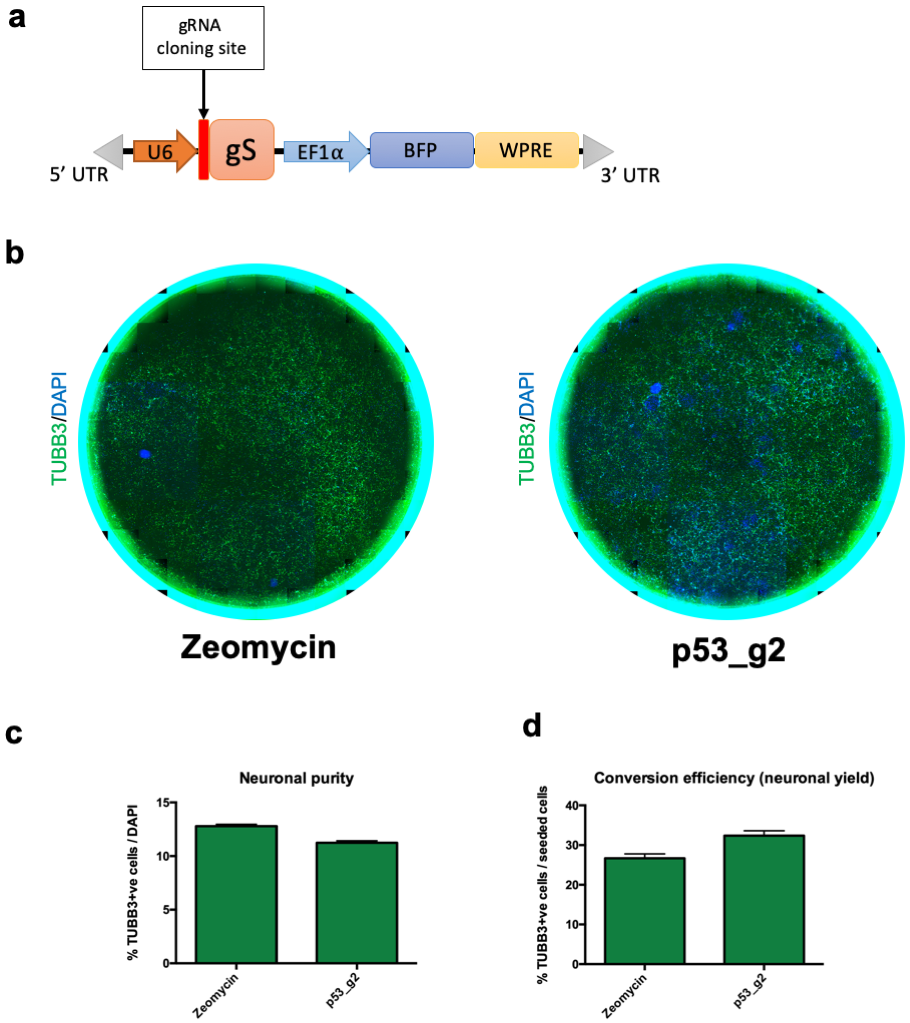


Fig. 3.2 – CRISPR/*Cas9* KO does not impact the conversion. (a) Schematic of the gRNA lentiviral construct. gRNA expression is driven by the U6 promoter. The expression of a BFP reporter within the same cassette is driven by the Ef1α promoter. gS: sgRNA scaffold sequence. WPRE: Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. **(b)** Whole well images of a conversion experiment. Cells were transfected with either the *Zeomycin* gRNA as a non-targeting control, or a previously tested gRNA against *p53*. **(b)** Quantification of neurons at day 15 of the conversion, shown as % of neuronal cells over total number of cells per field. 10 fields were analysed per replicate, n=2. **(c)** Quantification of the conversion efficiency towards neurons at day 15 of the conversion. The conversion efficiency

is calculated as the total number of neurons per well over the initial number of seeded MEFs and represents the neuronal yield of the conversion. 10 fields were analysed per replicate, n=2. Error bars indicate standard deviation.

3.2.2 Identifying a marker to selectively isolate iNs from the culture: testing available antibodies

The marker to isolate the induced neurons from the culture needed to be highly selective and suitable for flow cytometry. To start with, I made use of the available sc-RNAseq dataset (Treutlein *et al.*, 2016) to identify genes expressed exclusively in day 22 neurons. I then further reduced the list to proteins against which commercial antibodies were available. The mRNA expression of the candidate markers is shown in Fig.3.3a.

Antibodies against these candidate markers were first tested for their specificity by immunofluorescence on day 22 of the conversion (Fig.3.3b). All antibodies except the one against SYT1 specifically labelled neurons, as shown by the labelled cells co-expressing TUBB3. Antibodies against GPM6A, SNAP25 and SYP were chosen to be further tested in flow cytometry based on their good signal/background ratio.

Unfortunately, the antibody against GMP6A failed to identify a significant and distinct neuronal population via flow cytometry, as highlighted by the minor shift in the signal peak in Fig.3.4c. The signal was already not quite strong via IF, and, although flow cytometry is more sensitive in detecting low signal, it's possible that the antibody binding is suboptimal in suspension conditions and flow cytometry buffer.

The antibody against SNAP25 showed a more significant shift, though still failing to detect a separate population. The same shift was unfortunately detected in MEFs that had been subjected to the same conversion medium for 22 days, indicating that the signal detected by flow cytometry is non-specific (Fig.3.4d). A similar result was obtained using the antibody against SYP: although this antibody was able to identify a second, independent peak from the iN culture, the same peak was also identified in MEFs, suggesting again that the signal is non-specific. This high background detected via flow cytometry, and not detected via IF, is probably due to the higher

sensitivity and the issues it creates with antibody specificity in intracellular flow cytometry.

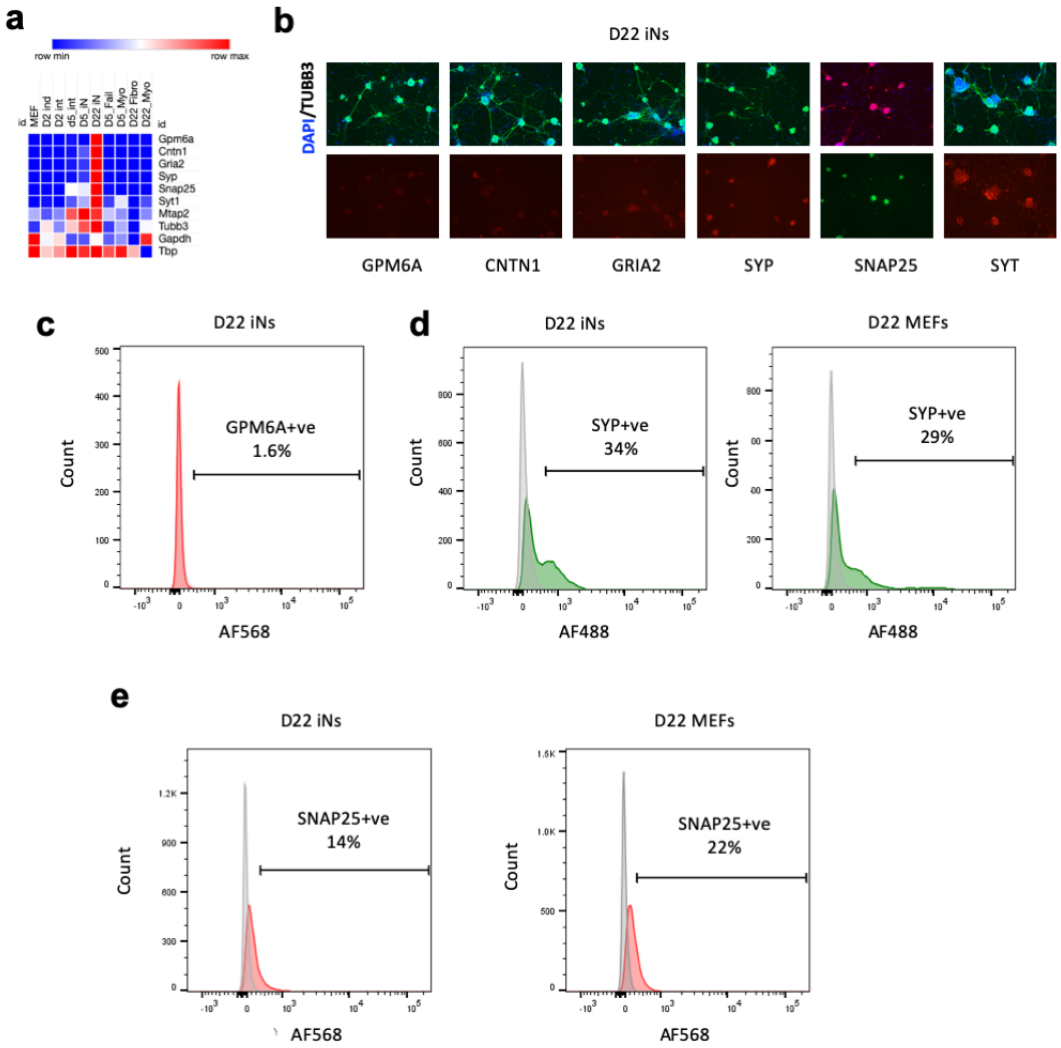


Fig. 3.3 – Isolating iNs using available antibodies. (a) Gene expression of the selected neuronal markers, *Tubb3*, *Map2* and the housekeeping genes *Gapdh* and *Tbp* – from Treutlein et al., 2016. **(b)** Immunofluorescence of iNs at day 22 of the conversion. **(c)** Flow cytometry profile of iNs at day 22 of the conversion, stained for GPM6A. The secondary antibody was conjugated to AlexaFluor568, the signal is shown on the horizontal axis. Signal in grey represents secondary-Ab only staining control. **(d)** Flow cytometry profile of iNs (left) and MEFs in TM medium (right) at day 22 of the conversion, stained for SYP. The secondary antibody was conjugated to AlexaFluor488, the signal is shown on the horizontal axis. Signal in grey represents secondary-Ab only staining control. **(e)** Flow cytometry profile of iNs (left)

and MEFs in TM medium (right) at day 22 of the conversion, stained for SNAP25. The secondary antibody was conjugated to AlexaFluor568, the signal is shown on the horizontal axis. Signal in grey represents secondary-Ab only staining control.

3.2.3 Identifying a marker to selectively isolate iNs from the culture: *hSyn1* promoter

Synapsin1 is expressed in neuronal cells of the CNS in humans. Its 470bp promoter has been shown to be selectively active and be able to drive reporter gene expression in *in vivo* and *ex vivo* mouse neurons compared to other cells of the CNS (McLean *et al.*, 2014). It has also been shown to be selectively active in neuronal cells derived by hESC reprogramming via *Ngn2*-overexpression (Patzke et al., 2015), and to mark the morphologically and functionally mature MEF-derived iNs (Black *et al.*, 2016; Adler *et al.*, 2012). Those two studies though did not investigate how broad is its expression within the conversion system, nor how early it can be detected. They also do not show any data regarding its use as a potential marker for isolation of neurons via flow cytometry.

I decided to address all these points and properly validate this construct for the purpose of iN isolation via FACS.

I obtained the *hSyn1-dsRed* construct from Addgene (Plasmid #22909). The *hSyn1-dsRed* expression cassette is already inserted into a pLV plasmid for packaging and transduction with lentiviral particles.

The diagram in Fig.3.4a summarizes the experimental procedure. Cells undergoing MEF to iN conversion were transduced at day 3 with lentiviral particles carrying the *hSyn1-dsRed* construct. By day 6 dsRED expression could be detected selectively from cells with pro-neuronal morphology. The expression levels increased progressively, as observed by the bright dsRED +ve cells with advanced neuronal morphology at day 15 of the conversion (Fig3.4b). At this point, dsRED expression was assessed with flow cytometry (Fig3.4c). dsRED expression levels presented as a broad spread rather than a distinct peak/cluster. This can be explained by two aspects:

1) The construct is inserted into the cells via lentiviral transduction, leading to different copy numbers and genomic locations of the construct, which will likely result in variable expression levels of *dsRed*.
2) The induced neurons are heterogeneous and likely have intrinsically different levels of *hSyn1* promoter activation

When non-converting MEFs were transduced with the same construct at day 3 and cultured under the same conditions as converting cells, very few cells show expression of dsRED at day 15. This shows that the *hSyn1* promoter has very little intrinsic leakiness and it is specific for neuronal cells in this context (Fig 3.4c, right panel).

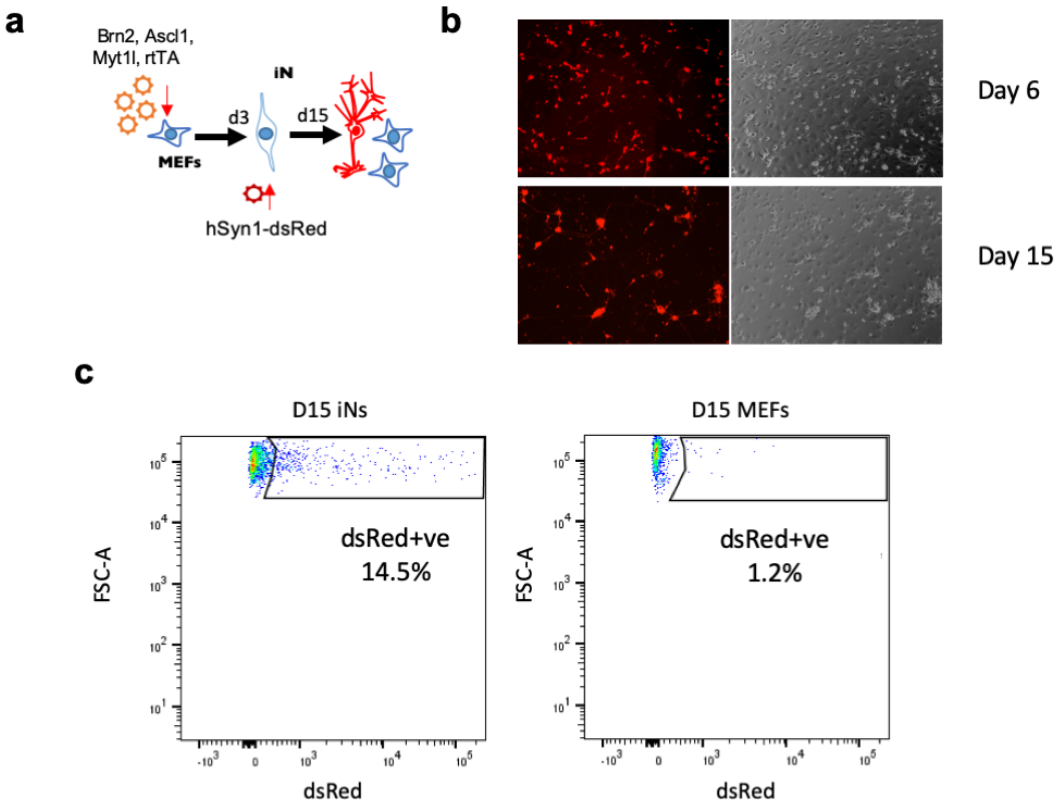


Fig. 3.4 – Isolating iNs using *hSyn1-dsRed* reporter activation. (a) Schematic diagram of a conversion experiment using the *hSyn1-dsRed* reporter. TF activation is induced in MEFs at day 0, one day post-transduction, as usual to initiate the conversion. At day 3, converting cells are transduced with lentiviral particles containing the *hSyn1-dsRed* construct. **(b)** dsRED

signal in cells with neuronal morphology is detected as early as day 6. By day 15 cells with clear neuronal morphology are showing selective *dsRed* expression. **(c)** Flow cytometry profiles of iNs (left) or MEFs in TM media (right) transduced with *hSyn1-dsRed* at day 3. A distinct population of dsRED +ve cells is only identified in the iN sample.

To verify that the *hSyn1-dsRed* expression is really being activated by neuronal reprogramming and not just the overexpression of the BAM factors, I looked into the activity of *Ascl1*, the pioneer master TF of the conversion (Wapinski *et al.*, 2013). *Ascl1* is a TF belonging to the basic helix-loop helix (bHLH) family and binds the E-box motif (Bertrand, Castro & Guillemot, 2002), preferentially the sequence CAGCTG (Wapinski *et al.*, 2013). This specific E-box motif is not contained in the *hSyn1* promoter, but the generic E-box motif CANNTG can be found twice within the promoter (as highlighted in Fig.3.5a). To exclude that dsRED expression is being driven simply by *Ascl1* binding on the *hSyn1* promoter, I converted MEFs into iNs using tetO-Ascl1 and hUb-rtTA only, and looked at *Ascl1* expression levels across the conversion using immunofluorescence (Fig.3.5b,c). At day 2 post-transduction, the majority (>90% cells) express relatively high levels of *Ascl1*. By day 15, cells that were transduced at day 3 with the *hSyn1-dsRed* construct and converted into induced neurons show neuronal morphology and dsRED expression. While many cells have downregulated *Ascl1* expression, there is a significant number of cells that is still expressing *Ascl1* at relatively high levels but did not activate dsRED expression. This suggests that *Ascl1* overexpression is not sufficient to activate the *hSyn1* promoter and that dsRED expression genuinely marks a population of converted neuronal cells.

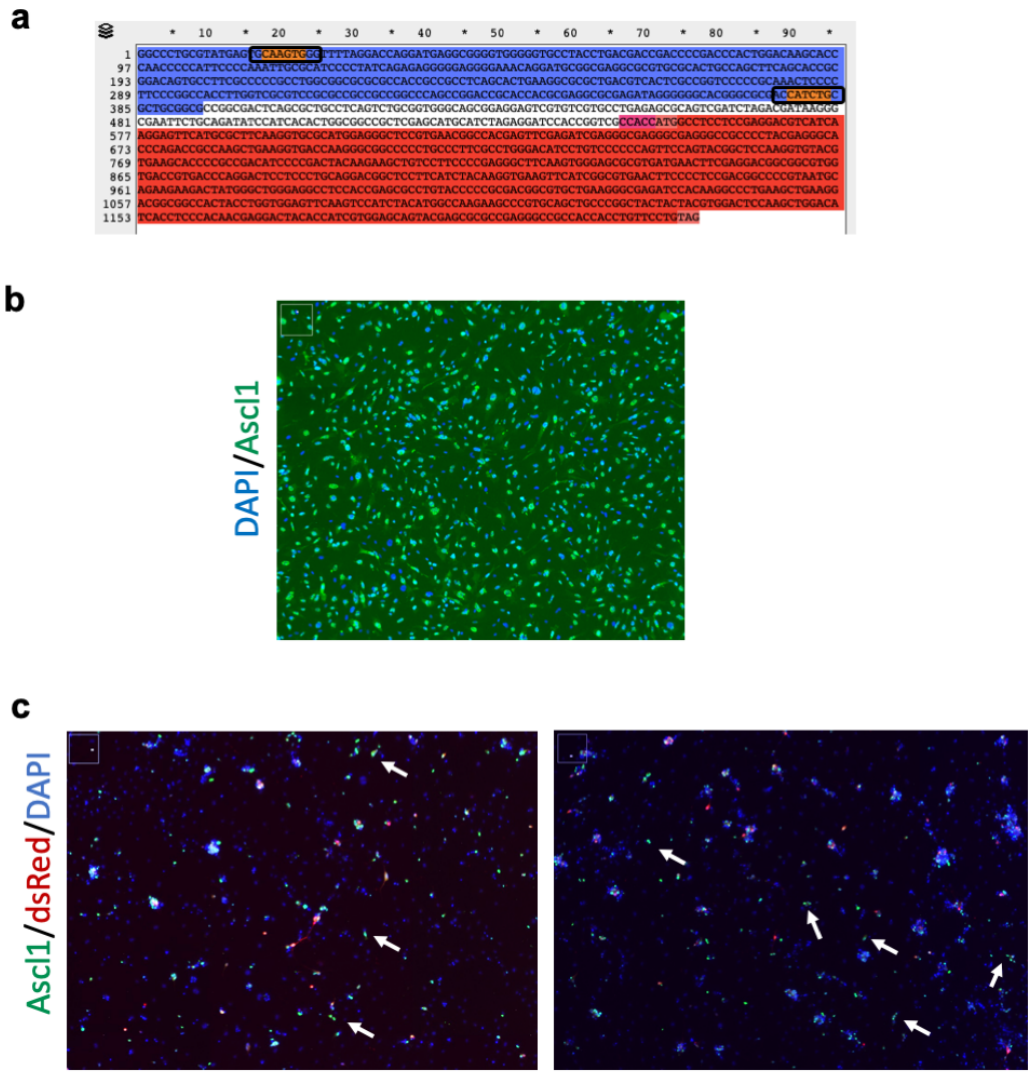


Fig. 3.5 – *Ascl1* overexpression alone is not driving *hSyn1-dsRed* reporter activation.

(a) *hSyn1-dsRed* construct sequence. *hSyn1* promoter in blue, *dsRed* cDNA sequence in red. Highlighted in orange are the identified E-Box motifs in the *hSyn1* promoter. **(b)** Immunofluorescence of *Ascl1* expressing cells at 48h post TF expression induction with doxycycline. >90% cells are expressing *Ascl1* (quantifications not shown) **(c)** Immunofluorescence of *Ascl1* and dsRED expressing cells at day 15 of the conversion (converting cells were transduced at day 3 with *hSyn1-dsRed*). White arrows indicate non-neuronal (dsRED -ve) cells expressing *Ascl1*. Cells with neuronal morphology are co-expressing *Ascl1* and dsRED, but a significant number of cells expressing highly *Ascl1* has not activated the *hSyn1-dsRed* reporter. The two fields are taken from two independent experiments.

To check that dsRED expressing cells are indeed induced neurons, I verified the co-expression of TUBB3 via immunofluorescence (Fig.3.6a). More than 95% of cells expressing dsRED also co-express TUBB3, as quantified in Fig.3.6b.

The dsRED+ve neurons were isolated by flow cytometry to assess the expression of a panel of neuronal and fibroblastic marker genes (Fig. 3.6c,d,e). dsRED +ve cells show higher expression of the neuronal markers *Gria2*, *Map2*, *Snap25*, *Stmn3* and *Tubb3* compared to dsRED -ve cells, indicating that dsRED +ve cells are of clear neuronal nature, though dsRED -ve cells are bound to contain neuronal cells that failed to activate dsRED, as well as cells that received the BAM factors and reprogrammed only partially, thus it is expected that they show expression of neuronal markers at higher levels than wt MEFs (Fig.3.6c). Compared to wt MEFs, both dsRED +ve and -ve cells show a decrease in expression of the fibroblast markers *Nestin* and *Fibronectin* (Fig.3.6d,e).

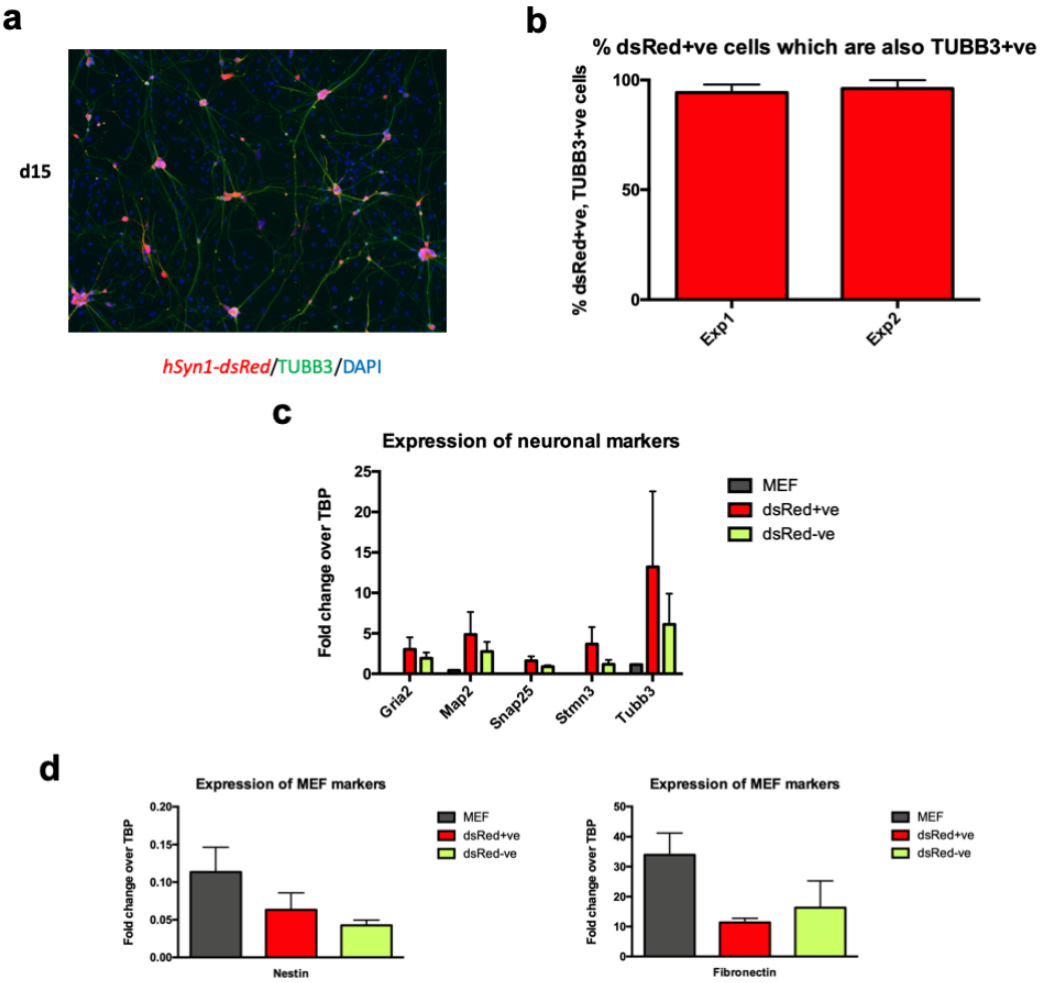


Fig. 3.6 –*hSyn1-dsRed* selectively marks neuronal cells. (a) Immunofluorescence for dsRED and TUBB3 expression at day 15 of the conversion. **(b)** Quantification of dsRED +ve cells expressing TUBB3. Results are the average of 10 independent fields in 2 biological replicates **(c)** Cells at day 15 of the conversion sorted for dsRED expression via flow cytometry express neuronal markers – as assayed by qRT-PCR – at higher levels than MEFs and dsRED -ve cells. Error bars represent STDEV, n=3. **(d)** Cells at day 15 of the conversion sorted for dsRED expression via flow cytometry have downregulated MEF markers *Nestin* and *Fibronectin*, as assayed by qPCR. Error bars represent STDEV, n=3.

To verify that the functionality of this construct is not affected by the CRISPR/*Cas9* KO system, I repeated the *p53* KO experiment shown in Fig.3.2 with cells transduced with *hSyn1-dsRed* at day 3. Cells with neuronal morphology and dsRED expression appeared by day 6 in both control and KO wells. The neuron purity in the culture, assessed by the % of dsRED cells at day 15 of the conversion shows no significant

difference from the control gRNA (Fig.3.8). This is consistent with the neuronal conversion efficiencies assessed by TUBB3+ve cells via immunofluorescence.

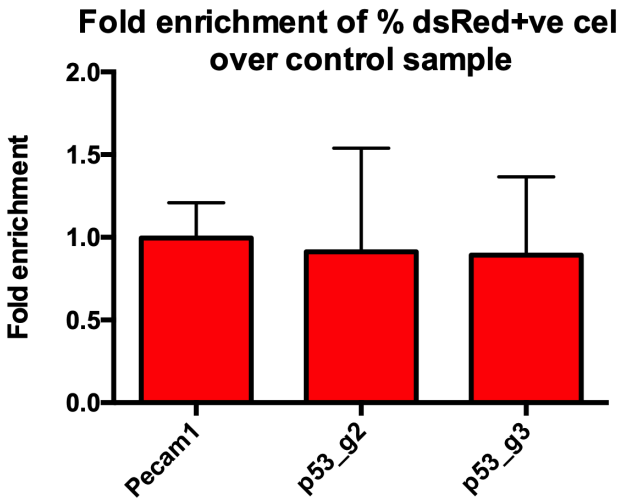


Fig. 3.7 – p53 KO phenotype investigated with hSyn1-dsRed construct. Day 15 flow cytometry profiles of iNs transduced with *hSyn1-dsRed* at day 3 and transduced at day -1 with gRNAs. The y axis represents fold enrichment over the % of iNs upon transduction with the *Zeomycin* non-targeting gRNA. *Pecam1* is included as an intronic targeting, neutral control (n=3). Error bars indicate standard deviation.

Overall, these data show that *hSyn1-dsRed* construct can be used to selectively label and isolate induced neurons from the remaining cell types in the conversion culture, and it is compatible with the CRISPR/*Cas9* KO system to be used for the screen.

3.2.4 Defining the screening conditions and selecting a library of genes to test

The next task was to identify the optimal screening conditions and calculate how many genes could actually be feasible to test in this set-up.

Based on previously published results from our collaborator Dr. Kosuke Yusa (Koike-Yusa *et al.*, 2014; Tzelepis *et al.*, 2016) and previous work performed in our lab (Kaemena, Beniazza et al., submitted for publication), we wanted each gene to be

covered by at least 5 gRNAs (to account for different editing efficiencies of the guides), and each gRNA ideally to be represented in at least 200 cells. The genome-wide gRNA library used in the papers mentioned above contains 90'000 gRNAs covering 18'000 genes in the mouse genome. To achieve the coverage desired, we would need to obtain and isolate 18 million induced neurons. Considering the average yield and purity of the conversion, processing these many cells through flow cytometry was not feasible. This led us to rule-out the genome-wide approach and decide to select a biased but smaller subset of genes to investigate.

To select the genes, I made use of the sc-RNAseq dataset analysing the iN conversion published in 2016 (Treutlein *et al.*, 2016). I selected differentially expressed genes (DEG) (>3 fold upregulated or downregulated) between the different populations of the conversion across time. I focused on genes that are likely to be involved in the conversion and maturation process of the induced neurons. Table 3.1 summarizes the cell types highlighted and how they were described/assigned in the original paper. Day 22 cells are classified based on their *Tau*-GFP reporter signal and the presence of marker expression (*Tnnc2* for myoblasts, *Eln* for fibroblasts) in the cells' transcriptomes.

Day	Label	Marker
day 0	MEF	-
day 2	induced	Ascl1 high
	intermediate	Ascl1 low
day 5	iN	Tau-eGFP +ve
	fail	Tau-eGFP -ve
day 22	iN	*Tau-eGFP* +ve, *Syp* +ve
	Myoblast	*Tau-eGFP* -ve, *Tnnc2* +ve
	Fibroblast	*Tau-eGFP* -ve, *Eln* +ve

Table 3.1 – Description of the cell types identified and described in the sc-RNAseq dataset.

The table 3.2 below lists the comparisons from which DEG were initially selected. The full final list of *1203* genes can be found in Appendix 1.

Nominator		Denominator	Criterion for choosing this comparison
d2 induced	vs	MEFs	early transcriptional response to Ascl1 induction
d2 induced	vs	d2 intermediate	high to low Ascl1 response
d5 iN	vs	d2 induced	activation of post-mitotic neuronal markers
d5 iN	vs	d5 fail	neuronal vs non neuronal early transcriptional response
d22 iN	vs	d5 iN	transcriptional differences between late and early iNs

Table 3.2 – Summary of the comparisons used to select the library genes. The column on the right describes the criterion used to select the specific comparison, i.e. I decided to select DEG between MEFs and day 2 induced cells to identify genes involved in the early response to BAM induction. Those are likely to be involved in downregulating MEF identity genes and upregulating neuronal reprogramming genes, and are thus interesting candidates to include within the screen, as they are likely to play a key role in cell identity switch.

Fig3.8 describes the biological and molecular role of the genes in the library as estimated by gene ontology enrichment analysis, performed via PANTHER (Anon, 2019; Mi *et al.*, 2019) .

There is a strong bias towards neural/neuronal function, but also broader terms indicating development, differentiation, cell communication and metabolism. These two main categories represent the two main aspects of the original 2016 study. The first 5 days of conversion focus on the cell identity change between MEFs and iN and include genes necessary to shut off the fibroblast identity as well as modify the metabolism requirements and membrane plasticity in order to morph and sustain neuronal activity. The second part focuses on the *in vitro* maturation of established iNs, which includes neurite growth and branching, development and maturation of

synapses as the neurons contact and form networks, and eventually the development of spontaneous firing (Vierbuchen *et al.*, 2010; Treutlein *et al.*, 2016).

From a molecular point of view, the majority of GO terms are related to signalling, receptor activity and membrane-bound proteins, which reflects the importance of these processes for neuronal development and function. Other terms include DNA binding/TF activity, indicating that many of the genes included in the library are important for modulating gene expression at a transcriptional level.

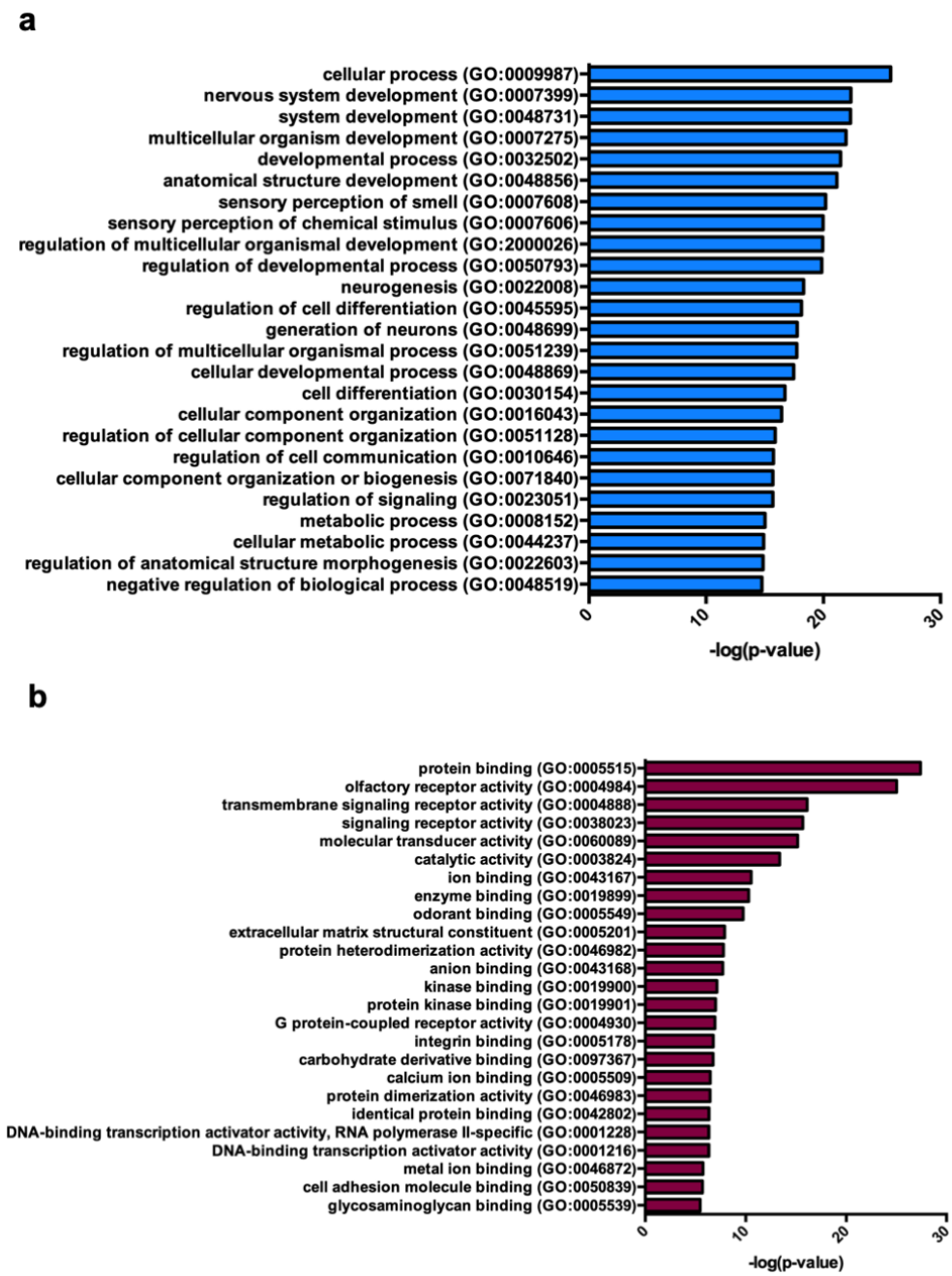


Fig. 3.8 –Gene Ontology enrichment analysis for the library genes. Analysis was performed on the *1203* genes using PANTHER. **(a)** Biological processes analysis, top 25 terms (out of 77). **(b)** Molecular function analysis, top 25 terms (out of 82).

3.2.5 Library cloning and quality check

The library of individual gRNAs was ordered as a single-strand DNA oligo pool synthesized by TWIST Bioscience. The gRNA sequences, 5 gRNAs per gene, were obtained from the mouse genome-wide library published by Dr. Kosuke Yusa's lab (Tzelepis *et al.*, 2016). To clone it into the pLV gRNA expression vector (from Fig.3.2), I followed previously published protocols, with minor edits (Koike-Yusa *et al.*, 2014).
The ss DNA oligo pool was made double stranded by PCR using primers annealing to the 5' and 3' shared sequences. The optimal number of cycles was determined by qPCR (Fig. 3.9a) to be 10. After PCR, a small aliquot of the product was run on a 2% agarose gel to confirm the presence of a single product at 79bp (Fig.3.9b). The PCR product was then column purified and digested with Bbs1 to isolate the gRNA-containing sequence to be ligated into the backbone. Due to the small size of the fragments after digestion, a PAGE gel was required to resolve the individual bands. The middle band corresponding to the 26bp containing the gRNA sequence pool was isolated and the DNA was extracted (Fig.3.9c).
Standard ligation was performed overnight into the pre-digested backbone and the resulting pool of plasmids was transformed into competent bacteria. The plasmid pool was then amplified, extracted and prepared for NGS.
To test our sequencing strategy, 3 technical replicates of the plasmid aliquot were prepared for sequencing, as outlined in the diagram of Fig.3.9d. The first step was to amplify the target sequence from the plasmid by PCR with primers containing overlapping ends to the Illumina Nextera kit. The number of cycles to be used was determined by PCR to be 16 (Fig.3.9e). The PCR product was run on an agarose gel to verify its size and purity, then extracted and eluted and re-verified (Fig.3.9e,f). It was then used to perform the second round of PCR, in which the i5 and i7 indexes from the Illumina Nextera kit were added to the final amplicon.

The individual libraries were purified with SPRI magnetic beads, pooled, and the quality of the pool was verified using TapeStation (Fig3.9g), then sent for sequencing on the Illumina Miseq.

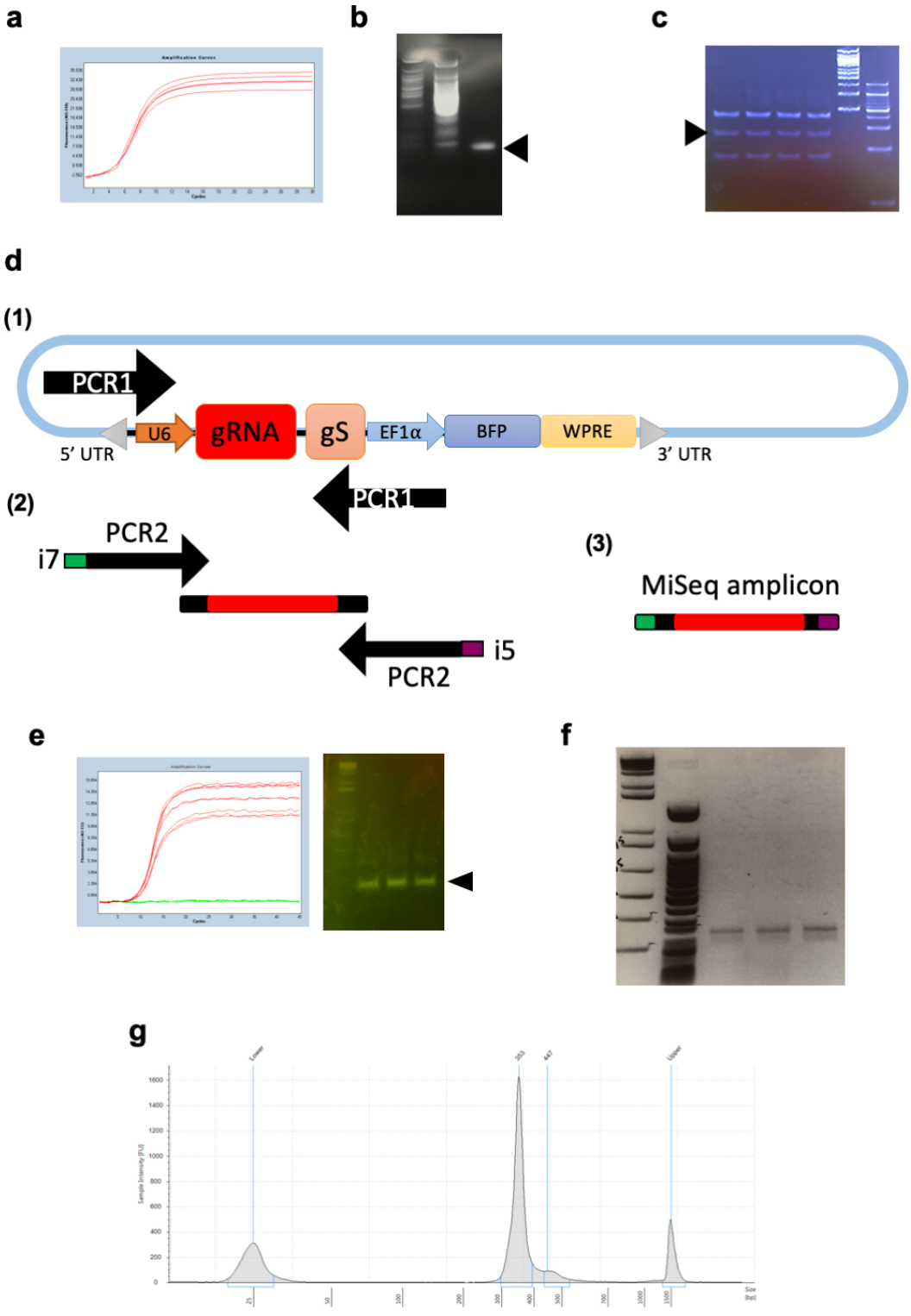

a
b
c
d
(1)
PCR1
U6
gRNA
gS
EF1α
BFP
WPRE
5' UTR
3' UTR
PCR1
(2)
PCR2
i7
i5
PCR2
(3)
MiSeq amplicon
e
f
g

Fig. 3.9 – Library cloning and preparation for NGS. (a) Amplification curves from qPCR. **(b)** PCR product from the 1st round of PCR run on a 2% agarose gel. The 2 left lanes are loaded with NEB 50bp ladder. **(c)** PAGE gel purification of the Bbs1-digested PCR amplicon pool from the first step of library cloning. The middle band at 26bp represents the correct fragment to be extracted and eluted. Ladders on the right are - left to right – NEB 50bp and ThermoFisher O'range 10bp ruler. **(d)** Amplification curves and agarose gel resolution of the 1st round of PCR for the preparation of NGS sequencing amplicons from the plasmid pool library. This fragment contains the gRNA sequences and 5' and 3' homology sequences to Illumina adaptors/barcodes. Ladder marker is NEB 1Kb+. **(e)** 2% agarose gel resolution of the second PCR product, after addition of Illumina adaptors. Ladder marker is NEB 1Kb+. **(f)** TapeStation results from the pooled library ready for Illumina Miseq.

All three technical replicates show a gRNA coverage, with the majority of gRNAs having around 250x coverage as estimated by the density distribution. The three replicates correlate well to each other as shown by the comparative distribution plots for P1vsP2, P2vsP3 and P1vsP3 (Fig.3.10a). When the replicates are normalized and averaged to combine gRNA read counts, the mode of the frequency distribution is 242, and the majority (96%) of gRNAs are represented within 10-fold difference in frequency, which is in line with what has been previously reported for similar gRNA libraries (Koike-Yusa *et al.*, 2014).

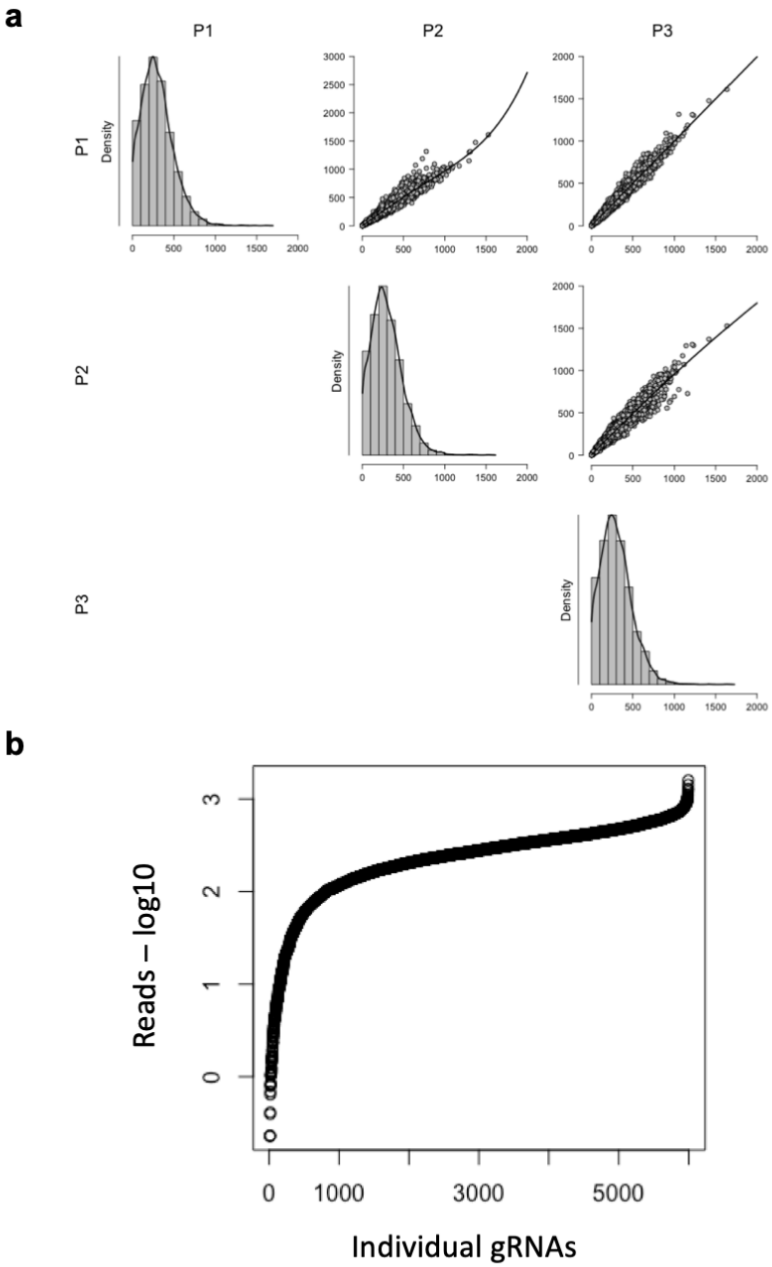


Fig. 3.10 – Library technical replicates sequenced with MiSeq show good library quality and gRNA coverage. (a) Density plots and correlation plots for each of the technical replicates show consistent average read count density and good correlation between each of the replicates. **(b)** log10 distribution of each gRNA counts – normalised and averaged from the three replicates. >95% of the gRNAs are within 10-fold of the average count density, consistent with previously published libraries.

These data overall indicate a good quality gRNA library ready to be packaged and transduced into cells.

3.3 Discussion

3.3.1 *hSyn1-dsRed* reporter marks iNs of all stages of maturation

The original iN conversion publication defines an induced neuron as cells with a circular soma with a neurite at least three times the length of their body (Vierbuchen *et al.*, 2010). This is because a few fibroblasts express *Tubb3* at low but detectable levels (as assessed by IF), and thus the presence of this marker is a necessary but not sufficient condition to identify a true neuronal cell. *Map2* expression is a more selective criteria when assessed by IF, though it is still present at low levels in non-neuronal fibroblastic cells (Vierbuchen *et al.*, 2010; Wernig *et al.*, 2002).

The *hSyn1* promoter has been previously used to mark directly reprogrammed iNs from human and mouse cells (Adler *et al.*, 2012; Black *et al.*, 2016; Drouin-Ouellet *et al.*, 2017), and has been described as targeting a subpopulation of neurons with more complex morphologies and more mature electrophysiological properties. Here I've reported the activity of *hSyn1* in driving *dsRed* expression in neuronal cells as early as day 6, when cells have still an immature neuronal morphology, with single or double neurite extensions, and have been reported to lack the ability to evoke action potentials and postsynaptic currents (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013).

In *vivo*, *Synapsin1* expression increases during neural development and neuronal differentiation, spiking during synaptogenesis. P5 mouse pups show about 25% of adult expression levels, that increase by day 10 to 50% and reach full expression at 1 month of age (Bogen *et al.*, 2009).

In vitro, endogenous *Syn1* is active in N2a mouse neuroblastoma cells, though its expression increases significantly upon differentiation and maturation of the cells into neurons (Paonessa *et al.*, 2013). *Syn1* is also expressed in human neuroblastoma cells, where its expression is limited by REST and is upregulated during neuronal differentiation (Lietz, Cicchetti & Thiel, 1998).

The previous studies using *hSyn1* in iNs do not mention how many of the iNs express *dsRed*, nor how early its expression can be detected. It stands to reason that these kinds of questions were not addressed because outside of the publications' scope, however it is possible that *hSyn1-dsRed* expression would have been detected from broader cell populations than reported. Adler and colleagues mention how *dsRed* expression was an excellent predictor of which cells would be able to produce evoked action potentials (24/26), but then describe how all other electrophysiological properties measured (resting membrane potential, action potential thresholds, amplitudes and numbers) were heterogeneous within those cells. Consistent with this, they mention how dsRED intensity correlated with the number of consecutive action potentials that the cells were able to fire. They also fail to show the electrophysiological properties of non-dsRED+ve iNs (Adler *et al.*, 2012).

It would be interesting to verify whether activation levels of the exogenous *hSyn1* promoter can be correlated to electrophysiological properties of the cells, i.e. if the intensity of the dsRED signal as assessed by FACS or quantitative fluorescence could be used to predict features like neuronal membrane potential or action potential evoked responses or the ability of the cells to spontaneously fire. Lentiviral transduction of the *hSyn1-dsRed* construct would create artefacts in this investigation, as integrated copy numbers and integration locus can potentially influence the levels of expression of *dsRed*. An engineered cell-line with the *hSyn1-dsRed* targeted within a safe, non-silenced locus would presumably lead to more consistent and homogeneous *dsRed* expression levels – or at least, the levels would not be dependent by lentiviral-derived artefacts, but from intrinsic heterogenicity of the iNs. Bright *dsRed* expression could be investigated as a potential reporter for iN maturation phenotype in a clearer and more quantifiable way.

3.3.2 *hSyn1-dsRed* reporter marks iNs with comparable efficiencies to *Tubb3* and other iN markers

In this thesis, I have been using the *hSyn1-dsRed* construct as a pan-neuronal marker after showing that it overlaps with virtually all *Tubb3* +ve neurons via IF. I've shown how the conversion efficiency and purity of the process in my hands are comparable to what has been previously reported, but it is worth discussing how efficiency and purity has been calculated in published works.

Tau, *Tubb3* and *Map* have been used to identify all MEF-derived iNs, regardless of the markers' molecular function, but differences in their roles are actually reported and suggest how certain conditions reveal marker promiscuity or conversely specificity to subpopulation of neuronal cells. For instance, while circa 95% of *Tau*-eGFP +ve cells are also *Tubb3*+ve when iNs are generated with the BAM cocktail, the percentage falls to less than 50% when iNs are derived from *Ascl1*-only reprogramming, suggesting *Tau* might be promiscuously marking cells with an induced neuronal fate regardless of whether they acquire subsequent *Tubb3* expression (Treutlein *et al.*, 2016). This was also characterized while studying *Myt1l* function: the DNA-binding region of *Myt1l* was fused to the repressor domain EnR, and then used to reprogram MEFs into iNs. Compared to wt *Myt1l*, this mutant form generates the same number of *Tau*-eGFP neurons, but only 30% of *Tubb3*+ve cells (Mall *et al.*, 2017). This highlights the importance of marker choice to characterise a phenotype, as in this example EnR-*Myt1l* is both enhancing (as assessed by *Tau*-eGFP) and not enhancing (as assessed by *Tubb3*) *Ascl1* reprogramming abilities.

Similarly, *Map2* seems to be a more stringent marker than *Tubb3*, as neuron counts from the same visual fields consistently report about 10% fewer *Map2*+ve cells than *Tubb3*+ve cells. (Wapinski *et al.*, 2013, 2017). Interestingly though, in *in vitro* cultured primary hippocampal neurons *Myt1l* knock-down leads to a decrease in protein levels of *Tubb3*, while *Map2* levels remain unchanged (Mall *et al.*, 2017). There seems to be a relatively unexplored relationship between *Myt1l* function and *Tubb3* expression that it's important to keep in mind when assessing neuronal conversions.

Assessing neuronal conversion purity with flow cytometry also requires particular attention. Keeping in mind the considerations about *Tau* mentioned above, sorting *Tau*-eGFP positive cells is likely to result in cells of neuronal nature but of heterogeneous maturity and marker expression, as discussed for *Tau*-eGFP sorted cells reprogrammed with *Ascl1* alone which also contain myoblast-like populations according on their gene expression profile (Treutlein *et al.*, 2016).

PSA-NCAM is also commonly used to mark fibroblast-derived iNs (Kim *et al.*, 2011; Barbagiovanni *et al.*, 2018). Barbagiovanni and colleagues reprogrammed iNs from MEFs using the BAM cocktail and report a neuronal purity at day 13 and 21 of the conversion of 60% and 70% respectively. Similarly, Adler and colleagues reported a 66% neuronal purity by day 12 as assessed by *hSyn1-dsRed* reporter activity (Adler *et al.*, 2012). Visual examination of immunofluorescence images from my work and other published works on MEF-derived iNs at day 13 and 21 suggests this might be partially an artefact, as by comparing DAPI nuclei counts to *Tubb3*+ve cells, the ratio does not reflect, and is in fact much lower than 60% to 70% (about 15% in this work, Fig.3.1 and Fig.3.4). This could be due to differences in cell death that would result in an enrichment of surviving neuronal cells compared to non-reprogrammed cells. It could also be the result of processing the cells for flow cytometry, as the iN clusters are much easier to detach from the cell culture dish compared to the fibroblastic cells that form sheet-like patches and require more mechanical effort to fully go in suspension. This also induces more cellular stress and causes a good portion of them to die off. I've found careful cell handling and consistent protocols for the cell lifting step to be crucial for consistent and reliable quantification of iNs via flow cytometry. This led me to have comparable, within the limits of replicate variability, neuronal quantifications between IF and flow cytometry. That being said, flow cytometry is a lot more sensitive, and considering the early activation *of hSyn1-dsRed* reported, it is likely that, like *Tau*-eGFP, some of the dsRed+ve iNs that I assess by flow cytometry would have incomplete reprogramming and heterogenous expression of *Tubb3* or *Map2*. It's important to highlight this and stress how identity and quantification of the iN conversion should be verified with more than one neuronal marker.

3.3.3 Library selection based on gene expression, gene networks or TF target sites

The library I designed contains strong biases towards neuronal genes. This may limit the discovery of novel genes or pathways involved in the conversion process, as many genes are already known to be involved in neuronal development. Moreover, expression levels and patterns are not always good predictors of the relevance of genes for the biological process of interest (Wapinski et al., 2017, Kaemena, Beniazza et al., submitted), so by selecting genes based solely on their expression, we are potentially missing out on a number of important genes.

On the other hand, the library bias towards neuronal genes will hopefully increase the likelihood that gene hits from the screen will have a broader biological relevance to neuronal development/differentiation and not just cell identity conversions.

Gene expression is just one of the ways to choose a subset of genes to investigate. Two recent publications studying *in vitro* neuronal differentiation and conversion decided to take a more functional approach instead and chose to screen for known or predicted TFs, reasoning that they'd be likely to be involved in cell fate decisions.

Liu and colleagues performed a CRISPR activation screen by designing 55,561 gRNAs targeting the promoters of all predicted TFs and DNA binding proteins based on an available dataset (Liu *et al.*, 2018). They performed the screen in mESCs, engineering the cell line for expression of the CRISPRa components, but also to engineering a *Tubb3* reporter system that allowed them to use MACS to sort out the neurons from the culture. They then validated their top hits on the MEF to iN conversion. Based on their gRNA ranking algorithm, they identify 74 genes as inducers of neuronal fate, and, importantly, 41 out of 74 showed the same expression levels in ESCs and neurons, confirming how change in gene expression is not a necessary condition for gene relevance or function. Out of their 14 top hits though, only *Ngn1* was able to generate *Map2*+ve *Tubb3*+ ve iNs by itself, whereas the presence of *Brn2*, *Ezh2* and *Foxo1* increased iN conversion efficiency by *Ngn1* significantly. They importantly showed how *Brn2*+*Ezh2* could generate neurons together, while neither can generate

neurons alone, and described how *Ezh2*, *Foxo1*, *Ngn1* and *Brn2* interact within the *Wnt/β-catenin* signalling pathway. *Ezh2* is part of the PCR2 complex and it is suggested to promote neuronal fate by repressing other identities. This highlights the strength of this TF oriented approach in making novel associations between genetic pathways and iN conversion.

Tsunemoto and colleagues also selected a list of 59 TFs (bHLH and POU containing) and arranged them into couples to generate a final library of 598 cDNA pairs that were cloned into dox-inducible vectors and delivered lentivirally into MEFs (Tsunemoto *et al.*, 2018). 13% of the cDNA pairs generated *Map2*/*Tubb3*+ve iNs. They then narrowed down their analysis on 5 combinations that included a TF that had not been associated with iNs before. The combinations include the factors *Neurog3*, *Ascl2*, *Brn3c*, *Neurod2*, *Atho1* and *Pit1*, all genes associated with neuronal function, but also Oct4 – one of the master regulators of pluripotency and particularly of induced pluripotency (Takahashi & Yamanaka, 2006). This is another example of how molecular function can be a good predictor of the gene's ability to reprogram iNs regardless of their expression or even biological relevance.

In terms of different library design strategies, it would have been an interesting alternative to perform a CRISPR/*Cas9* loss-of-function screen targeting TFs, as it would probably have a higher specificity for cell fate orchestrators and, because of the nature of loss-of-function screens, likely reveal roadblocks to cell fate plasticity. The intersection of the TF database (Liu *et al.*, 2018) and the CRISPR KO gRNA genome-wide library (Tzelepis *et al.*, 2016) would allow a straightforward selection of gRNAs without the need for *de-novo* design.

Making use of recently published datasets, it would have also been interesting to select TFs to screen for from the computationally generated TF network regulating iN generation and specification at day 5, 13 and 22 of BAM-mediated conversion (Wapinski *et al.*, 2017). The association of TFs with a specific time-point would make this network particularly suitable to investigate, with a loss-of-function screen, which TFs are actually required at each time point and how their presence affects neuronal induction and maturation.

Similarly, based on the work by Mall and colleagues on the newly discovered role of *Myt1l* as a suppressor of non-neuronal genes (Mall *et al.*, 2017), a loss-of-function screen targeting *Myt1l*-bound genes (from the ChIPseq dataset) would have good chance to reveal roadblocks genes that hinder the conversion towards iN by directing cells to alternative fates or resisting cell identity changes at all.

3.3.4 Library selection based on gene function and known pathways

A more biased approach to selecting a library would be to select genes within pathways that have been associated with direct neuronal reprogramming already. This is less likely to reveal novel mechanisms, but has higher chances of identifying hits. It would be a useful approach to characterise novel genes within the known pathways, or assign novel roles to known genes.

For instance, oxidative stress has been identified as a major barrier in direct neuronal reprogramming from both astrocytes and fibroblast (Gascón *et al.*, 2016). As ROS production had been already implicated in hematopoietic cell fate decisions (reviewed in Maryanovich and Gross, 2013) and neurons rely on different mechanisms for energy production compared to astrocytes and fibroblasts (oxidative metabolism vs anaerobic glycolysis typical of proliferative cells), Gascón and colleagues showed how between day 2 and day 4 of conversion, the majority of non-converting cells die off due to oxidative stress in the form of ferroptosis. The rapid switch to oxidative metabolism required by post-mitotic neurons leads to an accumulation of ROS and lipid oxidation, which most cells can't tolerate. This investigation revealed the essential role of *Bcl-2* in reducing ROS and promoting direct neuronal reprogramming. Thus, oxidative stress pathways and associated genes would represent a good subset to screen against in a loss-of-function genetic screen.

Recently the transcriptional repressor REST has been described as a neuronal reprogramming roadblock (Drouin-Ouellet *et al.*, 2017). This repressor works by

preventing gene expression of neuronal genes in non-neuronal cells, and the authors showed how REST knock-down significantly enhances reprogramming efficiency of adult human fibroblasts, partly through the upregulation of neural miRNAs miR-124 and miR-9, but also through upregulation of specific genes involved in synaptic formation and plasticity.

Thus, it is likely that REST targets could also represent reprogramming roadblocks, and it would be interesting to perform a loss-of-function screen against them to identify novel regulators and pathways that hinder this cell identity change.

Direct neuronal reprogramming has also been achieved for mouse and human cells by exposing fibroblasts to a cocktail of small molecules (Hu *et al.*, 2015; Li *et al.*, 2015). Li and colleagues showed how the combination of the small molecules Forskolin, ISX9, CHIR99021, and SB431542 acts in a synergistic way to generate, over 21 days, *Tubb3*+ve cells with neuronal morphology with a 30% culture purity. The substitution of SB431542 with I-BET151 led to 90% *Tubb3*+ve cells, but these cells were heterogeneous for other neuronal markers and had mixed subtype identity, with both VGLUT1- and GABA-positive neurons similarly represented. By selective removal of one molecule at a time, they identified ISX9 as the master inducer of neuronal identity, and I-BET151 as the master suppressor of fibroblast identity.

Hu and colleagues used a more complex combination of small molecules that included VPA , CHIR99021, Repsox, Forskolin, SP600125, GO6983 and Y-27632. In a similar fashion to the previous study, they identified CHIR99021 and SP600125 to be the ones responsible for neuronal fate induction, as their absence from the small molecule cocktail drastically reduces the neuronal conversion efficiency.

Characterising and understanding what the predicted targets of each of these molecules are would allow the identification of new pathways that control neuronal fate. Many of these compounds have quite broad roles though: for instance, SB431542 is an inhibitor of TGF-β signalling, a process involved very broadly in cell identity, differentiation and development (reviewed in Hata and Chen, 2016). It had already been associated with neuronal differentiation through the famous dual-

SMAD inhibition protocol developed to efficiently derive neurons *in vitro* from pluripotent stem cells (Chambers *et al.*, 2009). Similarly, CHIR99021 is an inhibitor of GSK3, part of the WNT signalling pathway that regulates development, but also adult stem cells maintenance and differentiation, and cancer (reviewed in Nusse and Clevers, 2017). It is famously part of the 2i conditions (2 inhibitors) that can maintain iPSC *in vitro* in the absence of LIF (Silva *et al.*, 2008).
It would be interesting to characterise the function of the target genes of these molecules to understand which actors within these broad and interconnected pathways come together and specify neuronal fate.

3.4 Summary

In this chapter I've set-up the experimental platforms and materials required to perform a CRISPR/*Cas9* KO screening in MEF-derived iNs.

Firstly, I verified the system's compatibility with CRISPR/*Cas9*-mediated KO. *Cas9*-expressing MEFs are capable of converting into iNs, and *Cas9* activity does not disrupt the conversion. *P53* knock-out leads to initial cell proliferation and increases total neuronal yields consistent with previously reported data.

Secondly, I characterised the *hSyn1* promoter as selectively active in iNs within the heterogeneous cell populations emerging from the conversion, as shown by observation of cells of neuronal morphology (live imaging), which co-express the pan-neuronal marker TUBB3 (immunofluorescence) and are selectively present in wells undergoing neuronal conversion and not in MEFs exposed to neuronal conversion medium (assessed by flow cytometry). dsRED+ve sorted cells express neuronal markers and downregulated fibroblast markers. The *hSyn1* promoter is not activated by *Ascl1* overexpression alone, but by the cell identity switch between MEF and iN overall.

This will allow the isolation of iNs from the culture by sorting dsRED+ve cells via flow cytometry, a step necessary when performing the screening in order to identify which gRNAs were integrated within the neuronal genomes and potentially impacted neuronal conversion efficiency.

As the conditions did not allow for an unbiased, genome-wide approach, I selected a subset of genes to screen against. The genes were chosen based on their expression in neuronal cells across the conversion, making use of the sc-RNAseq dataset (Treutlein *et al.*, 2016).
GO analysis revealed that these genes are mostly involved in neuronal identity and development, as expected based on the selection process, but also include broader functions regulating cell identity switches, cell differentiation and metabolism, and TFs regulating cell behaviour at a transcriptional level.
The final library consisted of a total of *1203* genes covered by 6000 gRNAs. The gRNAs were ordered as a pool of oligoes and cloned into the lentiviral backbone for transduction and U6-controlled expression. The quality of the library and the gRNA amplicon sequencing strategy were verified by NGS using MiSeq.

The next steps were to test the screening conditions in a small-scale pilot screen to verify whether the expected coverage was sufficient to pick up known phenotypes and then proceed with the screen. These points are the focus of the next chapter.

Chapter 4 Performing the CRISPR/Cas9 KO screen in MEF to iN conversion and analysing the results to identify hits

4.1 Introduction

Once the culture system and neuron reporter had been established, I decided to test the experimental conditions of the screen in a small-scale setting to check whether I could detect gRNA enrichment and depletion. I set up a pilot test using the KO of *Ascl1* as a candidate control to check for gRNA depletion: if a MEF is transduced with gRNAs against *Ascl1* it should not be able to be reprogrammed into a neuron because the gRNA should KO the exogenous *Ascl1,* which is part of the BAM cocktail (*Ascl1* is a single exon coded gene and the gRNA targets both endogenous and exogenous *Ascl1* copies).

Pilot testing was followed by performing the CRISPR/*Cas9* screen itself. The lentiviral gRNA pool was transduced together with the BAM converting factors. At day 15 post-induction of the BAM factors, induced neurons were identified and sorted using the *hSyn1-dsRed* construct. Both dsRED+ve and dsRED-ve populations were collected and analysed. gRNAs integrated within the cells were PCR amplified and sequenced to identify how each gRNA was represented within each population.

Sf3a1 was identified as an important gene in the conversion: its KO leads to a reduction in the efficiency of conversion. This gene is involved in pre-mRNA processing, and its role in the conversion is discussed in the final section of this chapter.

4.1.1 Aims of the chapter

The main aim of this chapter is to perform the screen and validate the screen result hits individually.

To achieve this, it was first necessary to try and verify that the screening conditions were appropriate to detect known phenotypes in a pilot experimental setup – which is the starting point of the chapter.
The second major part discusses the screen sequencing results and the different analysis used to interpret the data. A few individual gRNAs were selected from these results to be individually validated. This testing revealed that *Sf3a1* KO reduces the percentage of cells successfully able to convert to iNs.
The final part focuses on the validation of *Sf3a1* and some characterization of its role in the iN conversion.

4.1.2 Notes

1) gRNAs discussed in this chapter were selected from the previously published mouse genome-wide gRNA library that our collaborator Dr. Kosuke Yusa previously used to perform genome-wide screening (Tzelepis *et al.*, 2016).

2) NGS was performed in collaboration with EMBL Genomics Core Facility with the help of Dr. Vladimir Benes, Dr. Jelena Pistolic and Dr. Dinko Pavlinic.

3) NGS analysis through MAGeCK and data plotting was performed by Dr. James Ashmore.

4.2 Results

4.2.1 Pilot testing the screen conditions

To verify if the screening conditions were sufficient to detect enrichment/depletion of individual gRNAs, I first performed a pilot screen. The aim was to reproduce the screening conditions in small-scale and verify whether the planned coverage would

be sufficient to identify enrichment or depletion of gRNA sequences within neurons vs non-neuronal cells.

Considering neuronal yield based on the *hSyn1-dsRed* reporter discussed in the previous chapter I aimed to use, for the full-scale screen, 15 6-well plates per replicate to obtain a coverage to 450 iNs per gRNA. This upper limit was restricted due to the estimated processing time of cell preparation and FACS.

In the pilot test, I wanted to transduce 2250 starting MEFs per gRNA (resulting in 450 transduced iNs per gRNA based on 20% conversion efficiency) within one well of a 6-well plate and verify whether I could detect the depletion of *Ascl1* gRNA copies using qPCR. I chose to knock-out *Ascl1* because in the absence of *Ascl1*, *Brn2* and *Myt1l* can't reprogram MEFs into neurons (Vierbuchen *et al.*, 2010).

Lentiviral particles containing *Ascl1* gRNAs #1 and #2, a non-targeting control gRNA designed against the Zeomycin sequence and a targeting control gRNA against an intronic sequence of *Pecam1* were titrated so that I could infect 2250 MEFs of the 100'000 seeded in one 6-well. Both control gRNAs were predicted to not affect the conversion process and as such, should be present at similar levels in both *hSyn1-dsRed+* iNs and control MEF samples. At day -1, Cas9-MEFs received the BAM factors at MOI5 and the gRNAs mentioned above to the specific low titre. At day 0, dox was added to the medium to induce the expression of the BAM factors and start the conversion. At day 3, cells were transduced with *hSyn1-dsRed* to mark iNs. iNs at day 15 were sorted for dsRED expression, their genome extracted, and the presence of the gRNA within the genome was assessed by qPCR using the gRNA sequence itself as a reverse primer and an oligo annealing to the U6 common promoter as a universal primer for all the gRNAs being tested in the pilot (Fig. 4.1a). As a control sample I used non-converting MEFs transduced with the lentiviral particles carrying gRNAs at the same concentrations, but no BAM factors.

The expectations were that cells undergoing neuronal conversion that receive the gRNA against *Ascl1* – which also targets the exogenous copies of *Ascl1* transduced via lentivirus – would fail to become iNs. Thus, when analysing the genome contents of

iN, we should expect a depletion of the *Ascl1* gRNAs as compared to control transduced MEF. In these control MEF, *Ascl1* KO should not induce a loss of *Ascl1* gRNA from the genomic pool as it is not expressed in MEFs and thus is not likely to be required for their normal growth/survival.

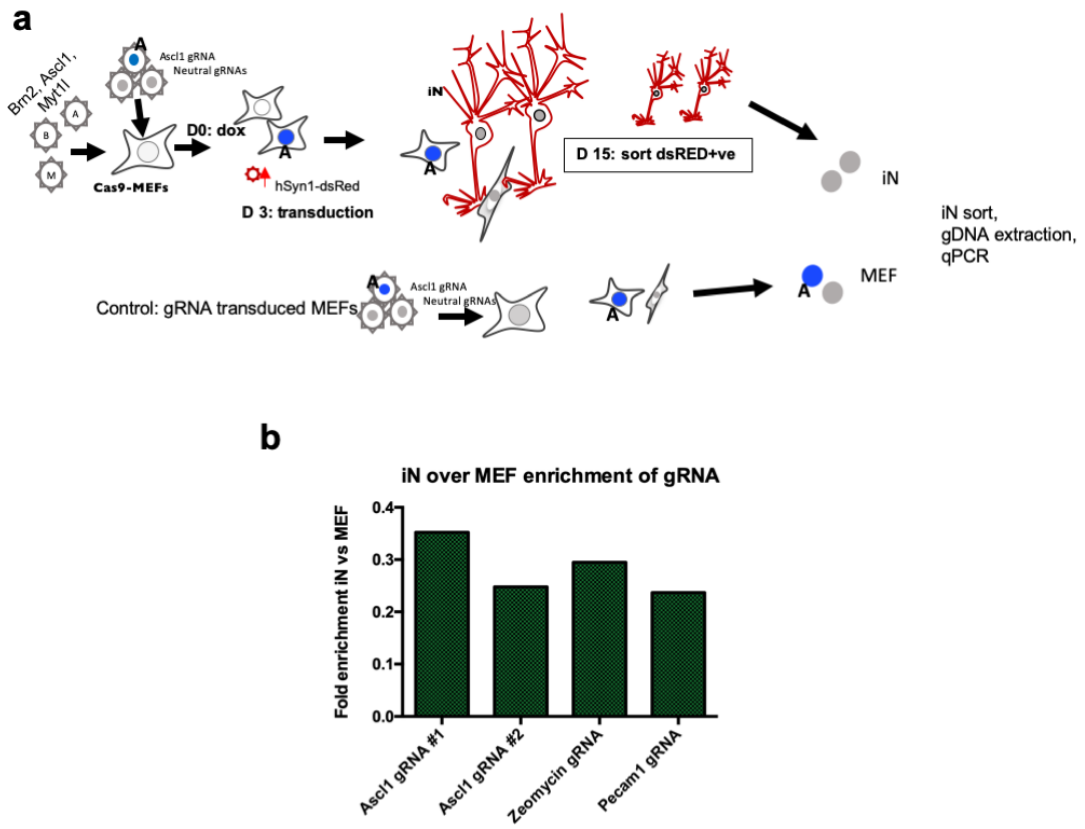


Fig. 4.1 – Pilot test to verify screen conditions. (a) Diagram of the pilot test. Transdifferentiating cells are transduced with gRNAs against *Ascl1* and Zeomycin (control) at the expected coverage that will be used for the screen. iNs are sorted, genome is extracted, and qPCR is used to assess the enrichment levels of each guide. As control, non-converting MEFs are transduced with the same gRNAs at the same coverage and analysed in the same way. **(b)** Pilot test results based on gRNA representation in each sample as assessed by qPCR. Y axis shows the fold enrichment of each gRNA between neurons (dsRED+ve cells) and non-converted MEFs.

The results of these pilot tests are shown in Fig.4.1b. When comparing the fold enrichment of each gRNA between iNs and MEFs, both *Ascl1* gRNAs are represented

in a comparable amount to the neutral control gRNAs Zeomycin and *Pecam1*. This is surprising as *Ascl1* gRNAs should not be present in iNs.

There are a number of possible explanations for this. The first thing I wanted to verify is whether the gRNAs against *Ascl1* are efficient in driving *Cas9*-mediated KO. I selected gRNA #1, the one that seems to be represented the highest, to test its indel and KO abilities.

I transduced *Cas9*-MEFs with *Ascl1* gRNA #1, together with tetO-*Ascl1* (and hUb-*rtTA*) at MOI5. I wanted to verify if, under optimal conditions for both gRNA expression and exogenous *Ascl1* expression, I could generate a KO phenotype that would result in the failure to reprogram MEFs into neurons.
Doxycycline was added 24h post-transduction as usual to induce *Ascl1* expression, and cells were allowed to reprogram for 15 days. At day 15 they were fixed and stained to check for TUBB3 expression. Results of this are shown in Fig. 4.2a: compared to a control sample where cells are transduced with the Zeomycin gRNA, when cells are transduced with *Ascl1* gRNA #1, very few neurons could be detected, indicating that under this condition this gRNA is able to knock-out exogenous *Ascl1* to impair the conversion.

Next, I wanted to test whether a single copy of the gRNA (which is representative of the pilot conditions) was able to induce indels efficiently. To achieve this, I transduced *Cas9*-MEFs with the gRNA at MOI=0.3, and then verified the presence of indels using TIDE (Brinkman et al., 2014). This algorithm uses chromatograms from sanger sequencing to identify the percentage of edited sequences compared to a reference, wild type amplicon, and quantifies the percentage of indels introduced by a specific gRNA.
I flow-sorted the cells based on expression of a BFP reporter contained within the gRNA lentiviral expression vector (see Fig.3.2 for more details) 72 hours post-transduction and extracted the genomic DNA. I then amplified a 500 bp region within

exon 1 of *Ascl1* from edited cells (and from non-transduced cells as a control) and sequenced this amplicon with Sanger sequencing. The chromatograms represent a pool of the genome editing occurring in each cell, so the TIDE software will use this information to align the edited sequences to the reference amplicon and assess the frequency of indels with a 1 bp resolution (https://tide.deskgen.com/ - Accessed on 15/04/18).

Results of this analysis are shown in Fig.4.3b: this gRNA, when transduced as a single copy within non converting cells, is able to induce indels with an efficiency of 92%.

The pilot conditions involved the gRNA being transduced as a single copy in cells with 7 overall copies of *Ascl1* coding sequence – 2 endogenous, 5 exogenous (most frequently, due to the lentiviral transduction of BAM factors with MOI5 each). To verify whether under these conditions the gRNA was still able to induce indels efficiently, I repeated the TIDE experiment, but transducing *Cas9*-MEFs with *Ascl1* gRNA #1 at MOI=0.3 and tetO-*Ascl1* (and hUb-*rtTA*) at MOI5, activated *Ascl1* expression 24h post-transduction with doxycycline and then harvested genomic DNA 48h later from BFP+ve sorted cells.

Results are shown in Fig.4.3c: under these conditions, the ability of this gRNA to induce indels is practically nullified, possibly because the number of copies of *Ascl1* to edit is too high.

This suggests that within the context of the pilot screen, the low copies of *Ascl1* gRNA would not have been enough to KO *Ascl1* and prevent MEF to iN conversion, thus it is reasonable to have detected the presence of this gRNA within neurons.

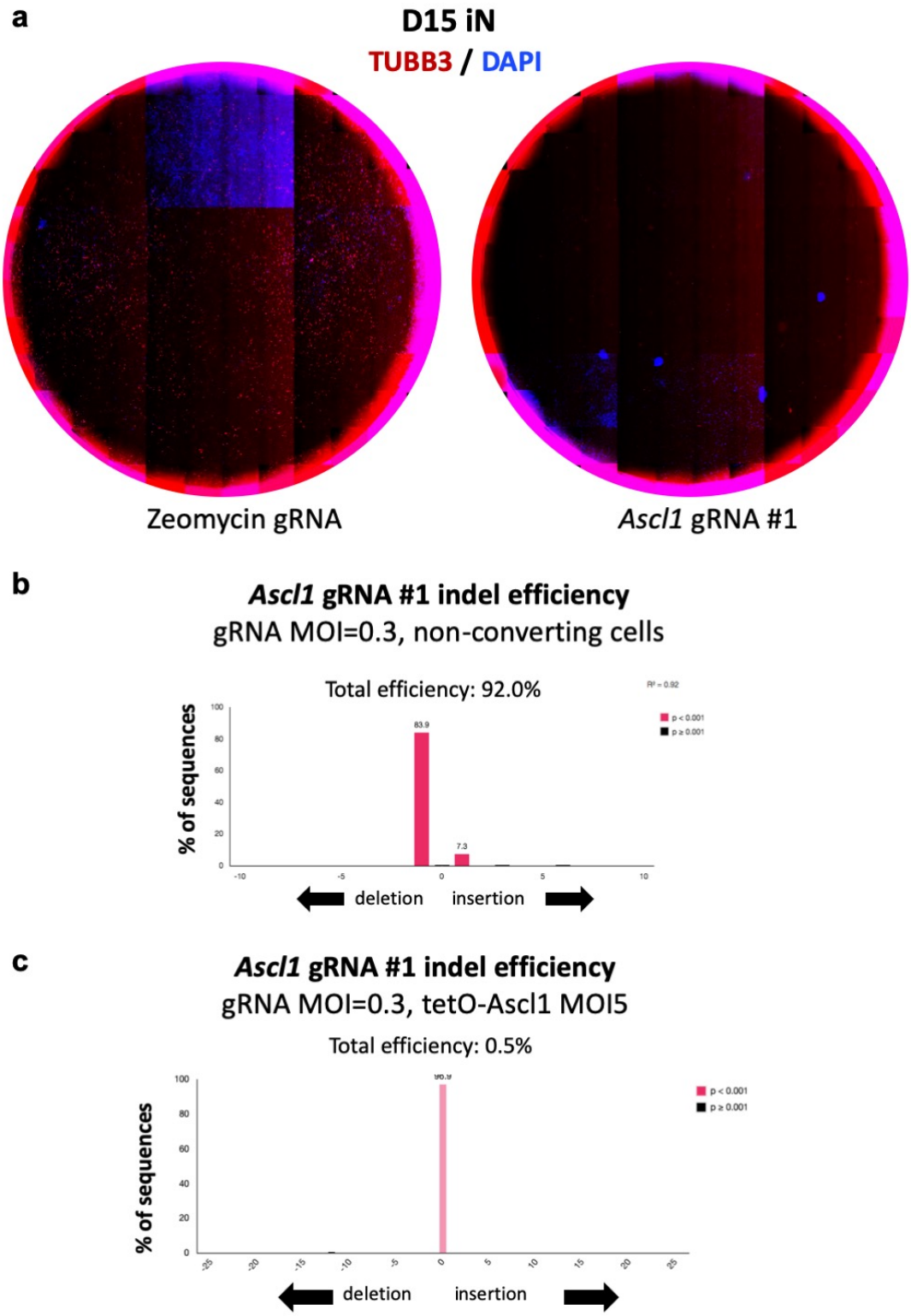


Fig. 4.2 – Ascl1 KO efficiency depends on Ascl1 copy number. (a) Day 15 iNs reprogrammed with *Ascl1* alone and stained for TUBB3 expression (in red), DAPI in blue to mark nuclei. Whole well images show strong signal when cells are transduced with Zeomycin gRNA (MOI5), but very few neurons when *Ascl1* gRNA #1 is transduced at MOI5. **(b)** TIDE

analysis for *Cas9*-MEFs transduced with *Ascl1* gRNA #1 at MOI=0.3, 72h after transduction. MOI=0.3 is sufficient to induce 92% indels at the start of *Ascl1* exon 1 (where the gRNA is targeting). **(c)** TIDE analysis for *Cas9*-MEFs transduced with *Ascl1* gRNA #1 at MOI=0,3 and tetO-*Ascl1* at MOI5, 72h after transduction. Indel efficiency is reduced to 0.5% when that many copies of *Ascl1* cDNA are present in the cell.

Overall this data suggests that using one of the reprogramming factors as a control for pilot testing was not ideal, as the insertion of multiple copies of the cDNA of the factor make it hard for CAS9 to edit them all when provided with a single gRNA copy.

I then went looking for a different candidate control gene that I could use in a pilot setting to test my strategy. I decided to test the pan-neuronal marker *Tubb3* itself, hypothesizing that cells unable to turn on this marker would fail to proceed with neuronal reprogramming. First, I needed to verify whether *Tubb3* KO would lead to reprogramming failure. I tested 5 gRNAs against *Tubb3* in an iN conversion experiment. gRNA #5 showed a clear KO phenotype as reported by immunofluorescence at day 15 of the conversion (Fig.4.3a). Generation of TUBB3+ve cells was almost completely blocked by the gRNA, indicating a complete KO at a protein level. MAP2 staining is still detected mainly in cell bodies, although at a lesser extent, and seems to indicate the presence of neurons.

Considering that the ultimate goal was to verify the presence of neurons with the *hSyn1-dsRed* reporter, I repeated the conversion experiment with *Tubb3* gRNAs #5 and transduced the cells with the *hSyn1-dsRed* at day 3. Cells were then analysed at day 15 with flow cytometry to quantify the presence of iNs based on dsRED expression (Fig.4.3b).

Compared to the control gRNA, *Tubb3* KO results in a 0.3 fold decrease in iNs (from about 30% to 20% dsRED+ve cells), indicating that the majority of cells lacking *Tubb3* are still functionally able to turn on *hSyn1-dsRed*, and thus *Tubb3* KO was not a suitable candidate for pilot-testing the screen.

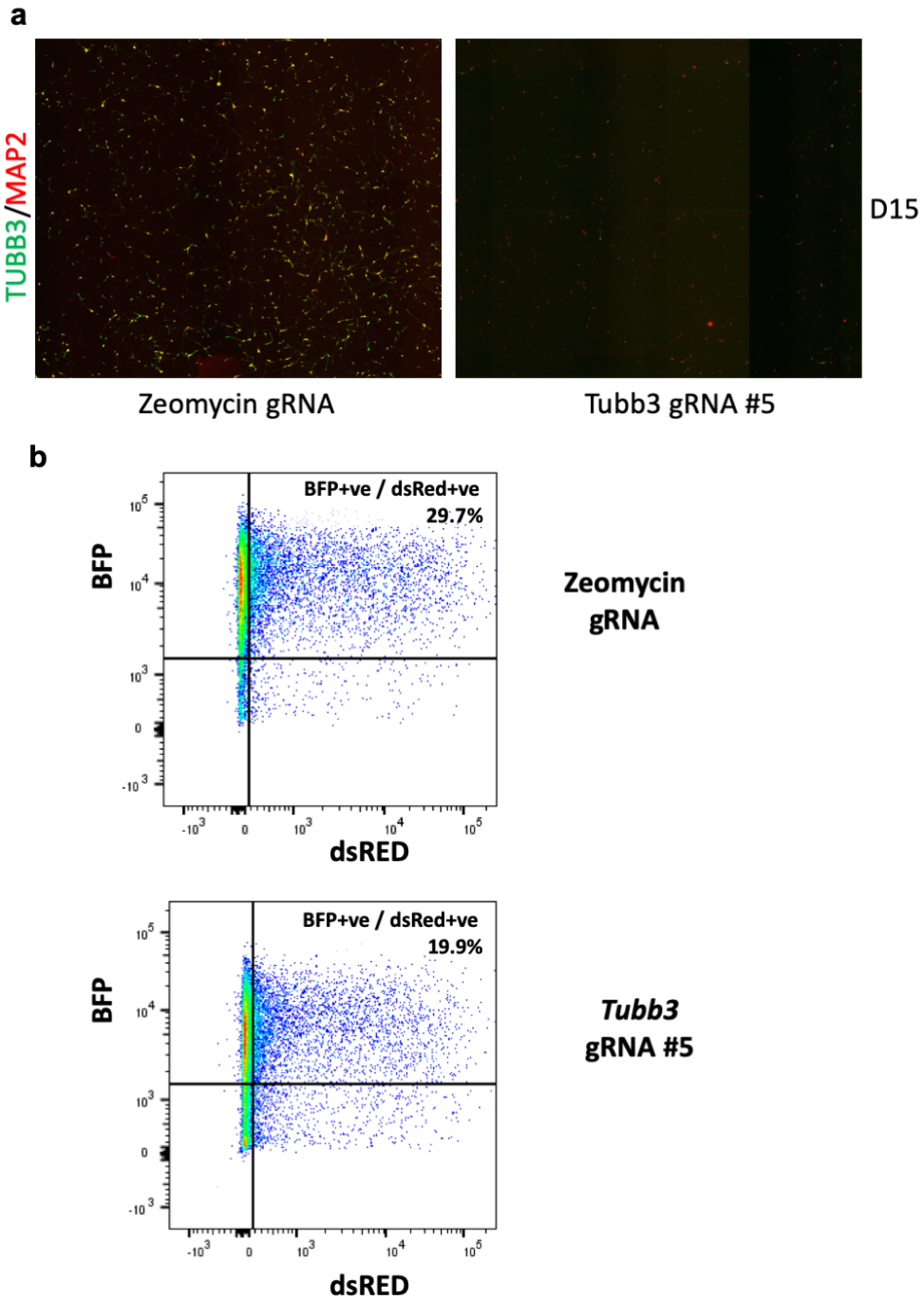


Fig. 4.3 – Tubb3 KO is unsuitable for pilot testing. (a) Day 15 iNs stained for expression of TUBB3/MAP2. Converting *Cas9*-MEFs were transduced with gRNA against Zeomycin (left) or *Tubb3* (gRNA #5, right) at MOI5. **(b)** Day 15 flow cytometry of iNs. Converting *Cas9*-MEFs were transduced with gRNA against Zeomycin or *Tubb3* at MOI5, and transduced at day 3 of the conversion with *hSyn1-dsRed*. Cells are gated on expression of dsRED (to mark iNs) and BFP (that is contained within the gRNA expression construct and marks cells transduced with the gRNA). Data represent one of 2 experiments.

4.2.2 Performing the screen

Being unable to find an appropriate control to perform more testing, I decided to proceed with the screen while maximising the coverage of iNs per gRNA, maintaining the estimated 450 iNs per gRNA.

A schematic of the screening process and FACS profiles for the screen are shown in Fig.4.4a.

I transduced a total of 9 million *Cas9*-MEFs per replicate with *Ascl1, Brn2, Myt1l, rtTA* at MOI5 together with the library virus pool at MOI3 – this should lead to an average of 3 gRNAs per cell and about 90% cells transduced with the library pool.

Doxycycline was added to induce BAM expression 24h after transduction (day 0). After 3 days, cells were transduced with *hSyn1-dsRed* construct to mark iNs. At day 15 cells were sorted based on their dsRED expression – both dsRED+ and dsRED- populations were collected and processed. The labelling of the samples in the next figures refers to “iNs” or “neurons” as dsRED+ cells, and “fibroblasts” as dsRED- cells, though this second population includes any cell that is non-neuronal at day 15, so fibroblast-like cells but also myoblast-like cells and cells of mixed identity based on what is known about the conversion from RNA-seq datasets.

Genomic DNA was extracted from these populations and libraries prepared by 2 consecutive rounds of PCR alongside 3 technical replicates prepared from the plasmid gRNA library into libraries to be sequenced in parallel. The 1st PCR amplified the gRNA from the genome or plasmid backbone in the case of the plasmid library. The purified PCR amplicon was then further amplified in a second PCR to attach Illumina sequencing adaptors and barcodes for multiplex sequencing. Library quality was verified with TapeStation, showing a single strong peak at the expected amplicon size (Fig.4.4b). The libraries were sequenced using the NextSeq platform.

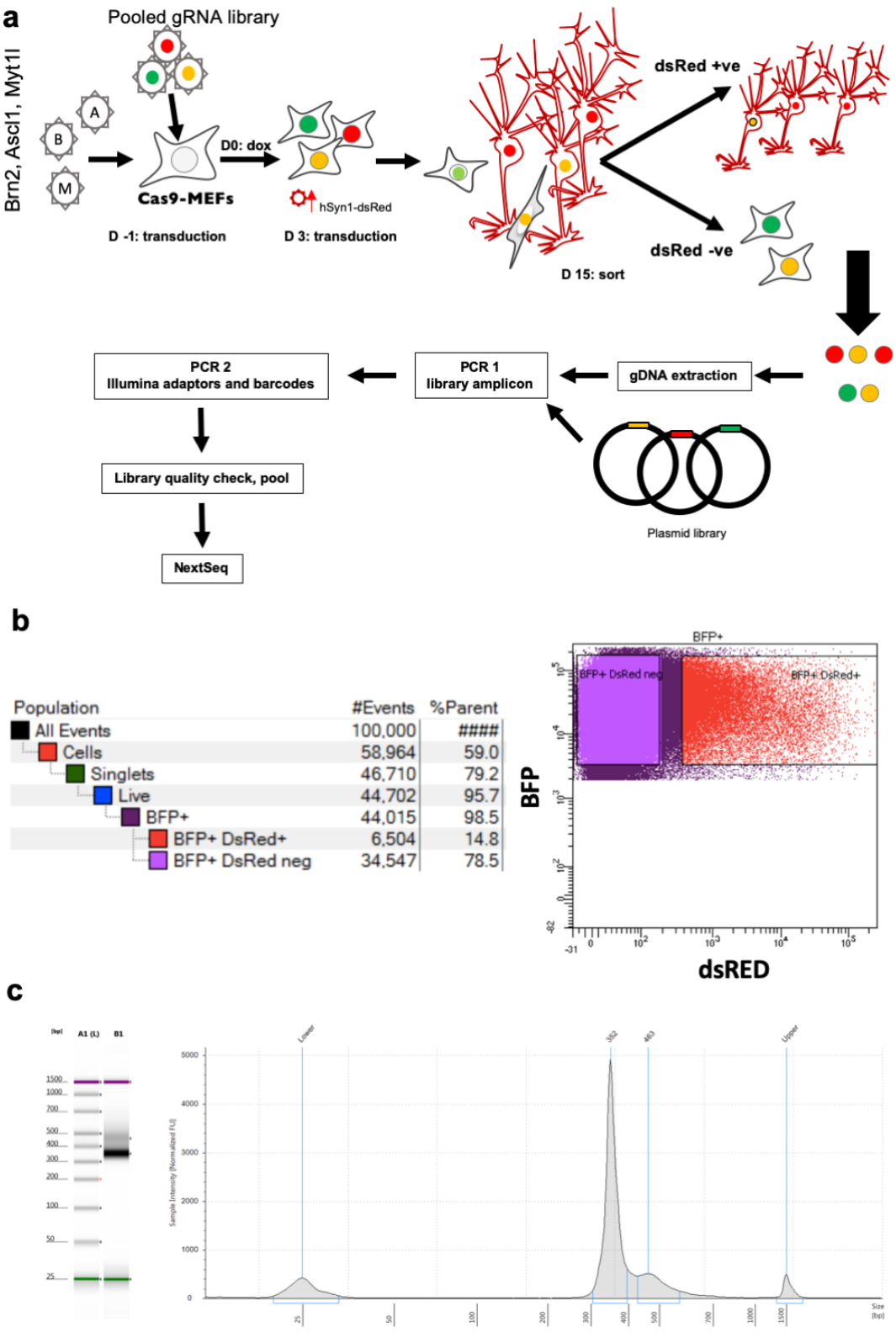


Fig. 4.4 – Screen setup and library preparation. (a) Explanatory diagram of the screen setup. *Cas9*-MEFs were transduced with the BAM factors and the library pool at day -1. Doxycycline was added the next day to induce BAM expression. At day 3, cells were

transduced with *hSyn1-dsRed* to mark iNs. At day 15 dsRED+ve and -ve cells were sorted, genome was extracted, and libraries were prepared by sequential PCRs, then purified, pooled and sent for NextSeq. **(b)**FACS gating strategy and plots for the final gate from one of the 3 screen replicates. Single live, BFP+ve cells were harvested based on dsRED signal within the gates shown on the right. The neuronal purity is estimated at 15%, but the gate for dsRED+ve cells was set to exclude cells with low-intensity signal. **(c)** TapeStation results of pooled libraries ready for sequencing. Left panel: TapeStation gel electrophoresis. Right Panel: chromatogram with flanking ladder markers. The small secondary peak at around 450bp is probably a result of concatamers, but the amount is quite small (<5%) and should not affect the amount of reads obtained from the main amplicon.

4.2.3 Screen results failed to show any significantly enriched or depleted gene

The sequencing results were analysed with the MAGeCK software. Briefly, this algorithm takes as input raw FASTAq files and outputs total read counts for each gRNA species present. It then checks within each gene how the 5 gRNAs have behaved in terms of enrichment or depletion between the test sample and the control sample, and based on their p-value and FDR assigns an FDR to each gene and determines if the genes present in the library show statistically significant enrichment or depletion.

Fig.4.5 shows the output table summaries for the comparisons between iN and plasmid library, iN and fibroblast, fibroblast and plasmid library. In all three cases, no genes were determined to be statistically significantly enriched or depleted (FDR <0.01 or 0.05).

Neuron vs plasmid comparison

	Comparison	Genes	Selection	FDR1%	FDR5%
1	Neuron1,Neuron2,Neuron3_vs_Plasmid1,Plasmid2,Plasmid3 neg.	1203	negative	0	0
2	Neuron1,Neuron2,Neuron3_vs_Plasmid1,Plasmid2,Plasmid3 pos.	1203	positive	0	0

Neuron vs fibroblast comparison

	Comparison	Genes	Selection	FDR1%	FDR5%
1	Neuron1,Neuron2,Neuron3_vs_Fibroblast1,Fibroblast2,Fibroblast3 neg.	1203	negative	0	0
2	Neuron1,Neuron2,Neuron3_vs_Fibroblast1,Fibroblast2,Fibroblast3 pos.	1203	positive	0	0

Fibroblast vs plasmid comparison

	Comparison	Genes	Selection	FDR1%	FDR5%
1	Fibroblast1,Fibroblast2,Fibroblast3_vs_Plasmid1,Plasmid2,Plasmid3 neg.	1203	negative	0	0
2	Fibroblast1,Fibroblast2,Fibroblast3_vs_Plasmid1,Plasmid2,Plasmid3 pos.	1203	positive	0	0

Fig. 4.5 – MAGeCK output tables show no significantly enriched or depleted gene in any comparisons. Output tables from MAGeCK for the neuron vs plasmid, neuron vs fibroblast and fibroblast vs plasmid comparison. Taking into account 3 replicates per sample, no gene had gRNAs with statistically consistent behaviour to result in enrichment or depletion with FDR <0.05.

This indicates that there is no gene that has gRNAs being enriched or depleted consistently and at a statistically significant level across the replicates. This could technically be due to a number of possible reasons. Firstly, it could be that KO of none of the genes targeted by the gRNA library leads to an enrichment or depletion of iNs numbers as detected by *hSyn-dsRed*. This is unlikely, considering the presence in the library, for instance, of *p53* – known to induce MEF proliferation upon KO, whose gRNAs should have been picked up as enriched at least in the comparison between fibroblasts and plasmid library (read count analysis for *p53* is discussed in Fig.4.8). Secondly, gRNAs targeting the same gene could induce opposite enrichment/depletion phenotypes, thereby confounding a statistical based analysis of hits. This is technically possible considering gRNA indel-inducing abilities might not necessarily result in nonsense-mediated decay, but could result in truncated proteins with different functions. However this is unlikely, based on previously published screen papers and data from the Kaji lab, that this would happen for many genes, and actually, data suggests in most cases that functional gRNAs that manage to induce indels generally lead to consistent phenotypes (Tzelepis *et al.*, 2016; Evers *et al.*, 2016). Thirdly, the KO phenotype responsible for the enrichment or depletion of gRNAs is too mild to be statistically significant at the coverage used for this screen.

At 48 hours post-induction of the BAM factors, all cells in the culture become post-mitotic, preventing a stronger difference between enrichment and depletion phenotypes being generated by the proliferation of successfully converting mutants. Based on the expected coverage, each gRNA species will have been transduced within 2250 MEFs and be represented in 450 iNs. A gRNA that decreases the conversion efficiency by 50% would instead be only carried in 225 iNs. However, these numbers are only averages; in reality the actual numbers of iNs with a certain gRNA will depend on experimental variation and the stochastic nature of the iN conversion, both factors that will introduce high background noise. Moreover, as cells die off due to the transduction and conversion stress, the bottleneck-effect will introduce artificial enrichment/depletion biases that will be specific for each replicate. It's possible that the phenotypes generated by the gRNAs are not strong enough to stand out over the technical variations in each experiment.

4.2.4 Neuron and fibroblast samples are significantly variable

To understand how variable the replicates of each sample are, we used gRNA read counts to perform Principal Component Analysis – results are shown in Fig.4.6.
The most striking feature is the separation of the neuron samples across both principal components, clearly indicating a high variability among all three of them. Fibroblast samples cluster closer, though there seems to be one of the three that is clearly separated from the others on PC1. Plasmid replicates instead cluster tightly close together. It's also important to notice how the fibroblast and plasmid clusters are very close together, and possibly explains why there's no differentially enriched gene when comparing those two samples.

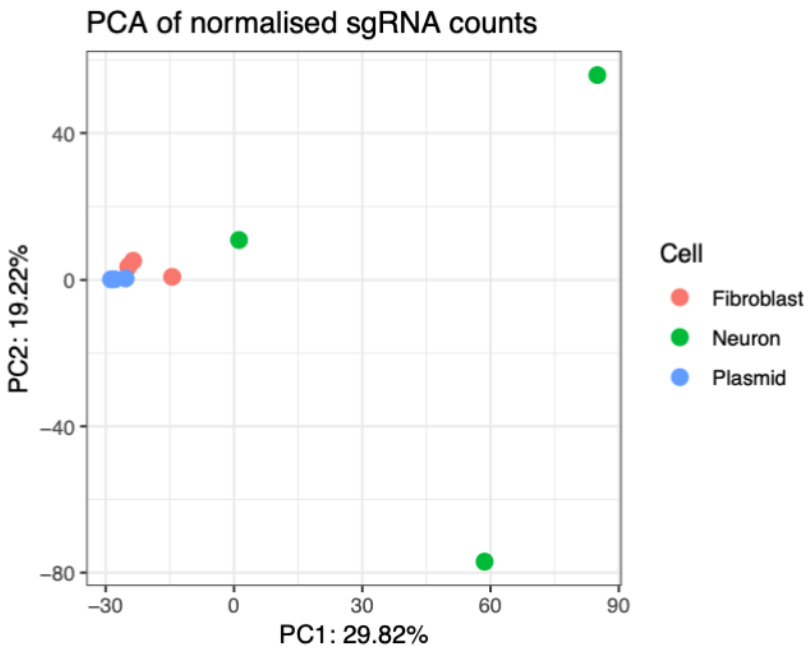


Fig. 4.6 – PCA of normalized sgRNA counts. gRNA normalised counts were used to generate a PCA plot to understand the samples similarities. PC1 shows a very significant separation between all three neuron replicates, and a smaller, but significant difference between one of the fibroblast replicates and the other two. PC2 separates again hugely the neuron replicates, and mildly the same fibroblast replicate again. Plasmid replicates cluster tightly together in both PCs.

The high variability between replicates will have made it difficult for any enrichment or depletion in gRNA to become statistically significant and offers one explanation as to why no gene was found to be enriched or depleted with a low enough FDR, at least when considering neuron samples. The variability is likely to be derived from the reprogramming process, which will select some gRNAs and exclude others regardless of gRNA function when cells die off due to the transduction and conversion stress. The BAM factors are also introduced lentivirally, which will result in heterogeneous levels of each of the factors within the neurons. This could be contributing for instance to some subset of neurons overcoming essential gene depletion thanks to higher expression levels of BAM.

Another thing to consider is that MEFs were transduced with the library lentiviral pool at MOI3, meaning that on average cells receive 3 different gRNAs. This was necessary because using an MOI0.3, which would have allowed to have only 1 gRNA per cell, would require transducing only 30% of the cells and thus would have required 3 times the number of cells to achieve the same library coverage.

This means that if one gRNA lead to an increase in iN conversion efficiency by disrupting a inhibitory gene, it would simultaneously enrich the other two gRNAs inserted within the same cell. Likewise, if one gRNA lead to a decrease in iNs by targeting an essential gene for the conversion, it would lead to an artificial depletion of the two gRNAs transduced in the same cell. However, since the chances of the same 2 gRNAs being co-transduced with that gRNA are extremely low, this should not be a major concern in terms of generating false positives or false negatives but it might contribute to increase background noise in gRNA read counts and, since it will be specific for each replicate, will lead to increased variability.

4.2.5 gRNAs within cells are not exposed to high selective pressure

The next thing I wanted to address was the representation of gRNAs within each comparison to understand the distribution of the enrichment and depletion of gRNAs. We used MA plots to show the fold enrichment over the read counts of each gRNA for the neuron vs plasmid comparison, neuron vs fibroblast comparison, and fibroblast vs plasmid comparison. Plots are shown in Fig.4.7.

The three comparisons have relatively similar distributions, with the first plot showing a slight trend towards a linear relationship between fold change and average gRNA count (the cluster of gRNAs is slightly oriented towards a diagonal line). Two observations can be made, that very few gRNAs show greater than two-fold enrichment (> 1 $\log_2$ [fold change]) and that gRNAs with high $\log_2$(fold change) have very low read count. gRNAs with low read counts are bound to have higher technical variance and won't achieve statistical significance.

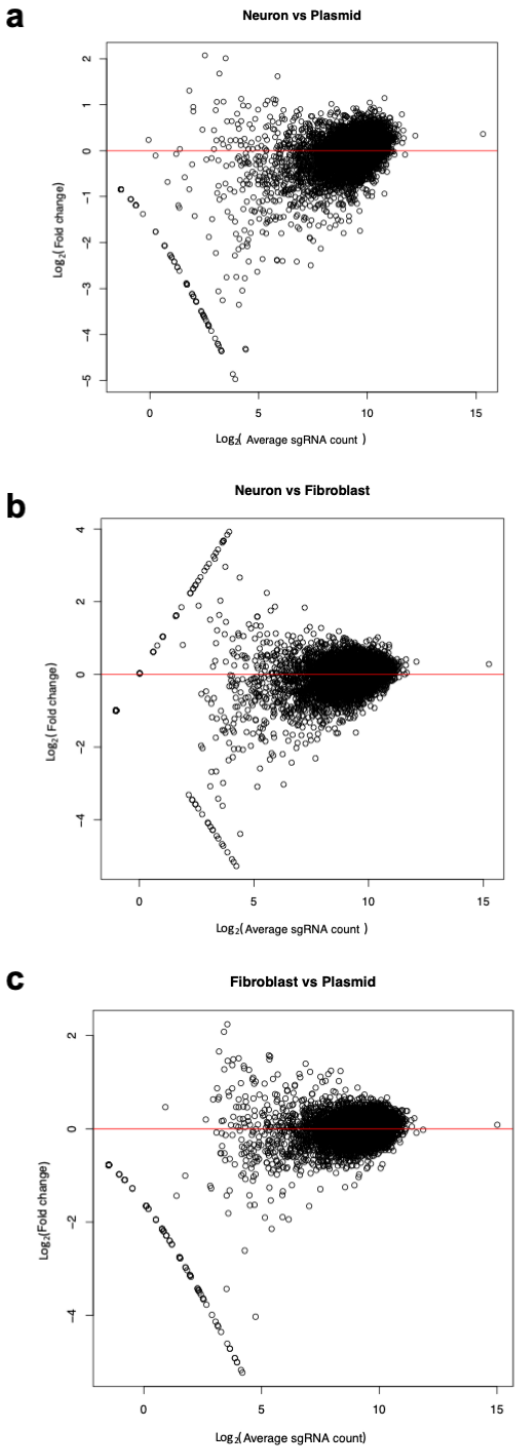


Fig. 4.7 – MA plots for each comparison. Plots are comparing $\log_2$(fold change) between the normalised replicate averages of each sample against the $\log_2$ of the average sgRNA count within the samples. **(a)** Neuron vs plasmid comparison. **(b)** Neuron vs fibroblast comparison. **(c)** Fibroblast vs plasmid comparison.

To explain this, it is helpful to consider the "life-span" of gRNAs during the conversion. The only criterion for selective pressure against gRNAs in this screen is whether their presence will allow a cell to convert into a neuron or not. A gRNA targeting a potential roadblock will increases the chances of MEFs becoming neurons, so should be enriched in the neuronal population. However, because neurons are post-mitotic, the copy number difference compared to a neutral gRNA which does not affect the conversion would not get amplified by cell proliferation – a phenomenon that other screens instead rely on, for instance screens looking at genes involved in cell survival/growth, or reprogramming to a proliferative cell type like iPSCs (Koike-Yusa *et al.*, 2014; Shalem *et al.*, 2014; Zhou *et al.*, 2014; Tzelepis *et al.*, 2016; Kaemena, Beniazza et al., submitted for publication).

Conversely, a gRNA targeting a potential essential gene for the conversion will prevent a cell from becoming a neuron, and is supposed to be underrepresented in the neuronal population. However, as the rest of the cells also stop proliferating, there is not a lot of chance for it to be out-grown.

The strength of the phenotype of each gRNA relies on its efficiency to induce indels at single copy. The biological result of the KO – abolishing iN conversion, would need to be particularly strong (e.g. result in all cells that receive the gRNA to not convert into neurons) to be picked up as a significant decrease in gRNA read counts in the neuron population.

To try and understand better individual gRNA behaviour, reported in Fig.4.8 are the normalised read counts for all 5 gRNAs targeting *Ascl1* and *p53*. Looking at *Ascl1* (top table) we can notice straight away how gRNA counts are a lot closer in plasmid replicates compared to fibroblasts and neurons. gRNA #1 actually does show lower read counts in neuron samples compared to plasmid replicates – but the difference between neurons and plasmid means is actually a 0.7 fold decrease, which is quite minor. All the other gRNAs show very little to no decrease in read counts, consistent with inefficient KO of exogenous *Ascl1*.

Looking at *p53* instead, it's interesting to focus on plasmid and fibroblast replicates. When *p53* gRNAs are transduced at MOI5, we observe significant proliferation of cells within the first 48h. Looking at the read counts from the screen, a net enrichment in the fibroblast samples can only be identified for gRNAs #3 and #4, but again, the fold-change is well below 1. This could be due to the fact that cell proliferation that is restricted to the first 48h is not enough to allow gRNA read count inflation. It could also be due to too low coverage: too few cells receive *p53* gRNAs and their proliferation is not enough to be picked up from the noise.

sgRNA	Gene	Plasmid1	Plasmid2	Plasmid3	Fibroblast1	Fibroblast2	Fibroblast3	Neuron1	Neuron2	Neuron3
Ascl1_g1	Ascl1	475.6	501.8	508.7	491.5	227.6	519.6	388.5	213.1	348.0
Ascl1_g2	Ascl1	660.6	620.5	647.6	709.0	807.2	670.0	690.2	888.7	763.8
Ascl1_g3	Ascl1	788.6	748.2	804.4	530.8	708.5	659.0	777.8	667.3	652.9
Ascl1_g4	Ascl1	1016.3	1018.4	983.9	949.9	1018.6	1210.7	1069.5	1276.4	1305.7
Ascl1_g5	Ascl1	368.9	372.6	337.6	313.3	322.9	285.9	205.0	324.3	296.4

sgRNA	Gene	Plasmid1	Plasmid2	Plasmid3	Fibroblast1	Fibroblast2	Fibroblast3	Neuron1	Neuron2	Neuron3
Trp53_g1	Trp53	280.5	327.0	306.4	299.8	415.8	183.6	194.2	376.3	297.2
Trp53_g2	Trp53	257.1	224.0	241.8	308.4	184.7	180.5	219.0	69.6	199.0
Trp53_g3	Trp53	796.8	847.5	772.1	800.0	1147.5	1082.3	667.8	1327.3	1068.6
Trp53_g4	Trp53	509.2	504.0	468.0	779.1	720.1	689.1	910.8	1256.7	689.3
Trp53_g5	Trp53	367.9	330.0	363.9	403.1	279.9	373.1	426.5	179.8	226.1

Fig. 4.8 – Normalized read counts for Ascl1 and p53 gRNAs. Plasmid replicates are in grey, fibroblast replicates in blue, neuron replicates in red. Upper table: *Ascl1* gRNAs read counts for all three replicates of each sample. Lower table: *p53* gRNAs read counts for all three replicates of each sample.

Overall this data suggests that the screen conditions were not powerful enough to pick up on phenotypes. This is likely because of the high variation within replicates, and because phenotypes are not strong enough to induce a big fold-change.

4.2.6 Selecting 22 gRNAs to test individually from the screen results

Because there were no significantly enriched or depleted genes, I decided to look at the behaviour of each gRNA and select a few to test individually and at high MOI. I

used FDR as a selection criterion, regardless of whether it was calculated against enrichment or depletion. The assumption was that the gRNAs that managed to obtain a low FDR had better chances of being the result of a real phenotype.

Fig. 4.9 summarises the selection process. I selected, from the neuron over fibroblast comparison, the gRNAs with low FDR (for either enrichment or depletion). I then did the same for the neuron over plasmid comparison, and then overlapped the two lists. This way, I identified 10 gRNAs from 10 different genes, listed in the box on the right (Fig.4.9a). Moreover, I also used gRNAs with FDR <0.05 from the neuron over fibroblast comparison to select a set of genes that had 2 or more gRNAs within this list (Fig. 4.9b). I then tested the two gRNAs present in the list.

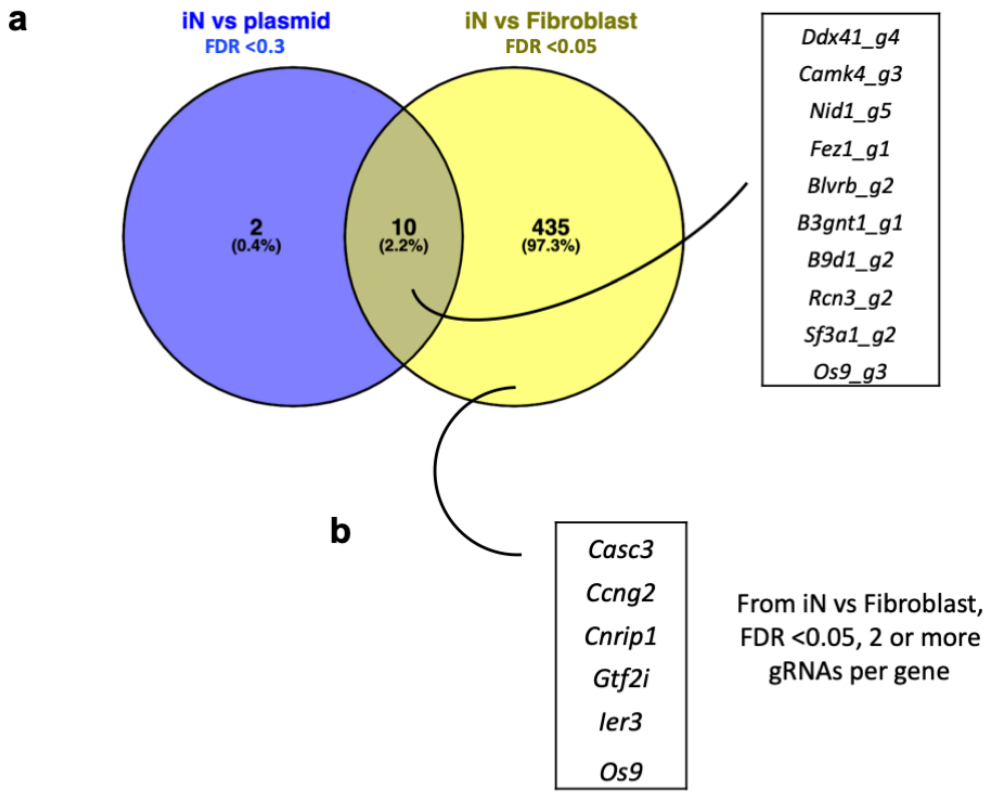


Fig. 4.9 – Selection of gRNAs to test individually from the screen results. (a) gRNAs from the iN vs plasmid comparison with FDR <0.3 (regardless of enrichment or depletion direction) were intersected with gRNAs from the iN vs Fibroblast comparison with FDR <0.05. The 10 gRNAs resulting from the overlap are listed in the box on the right. Graph generated with Venny (Oliveros, 2007). **(b)** Genes from the box were selected from the IN vs Fibroblast

comparison if they had 2 or more gRNAs with FDR <0.05 targeting them. The 2 gRNAs identified were tested individually. *Os9* appeared in both lists, gRNAs #2 and #3 were tested.

gRNAs were cloned into the pLV-U6-gRNA-BFP vector as previously described, packaged into lentiviruses and transduced into converting *Cas9*-MEFs at MOI5. Converting cells at day 3 were transduced with *hSyn1-dsRed* to allow quantification of neurons at day 15 via flow cytometry. The results of testing the above gRNAs are shown in Fig. 4.10. Virtually all gRNAs failed to show a clear phenotype, as the fold enrichment compared to the non-targeting Zeomycin control is around 1. The notable exception is gRNA #2 targeting *Sf3a1*, which results in a 50% decrease in conversion efficiency.

To verify that this phenotype was due to *Sf3a1* KO and not off-target effects of the gRNA, I tested two more guides from the same gene (Fig. 4.10b). All three gRNAs consistently reported a 50% reduction in neurons at day 15 of the conversion.

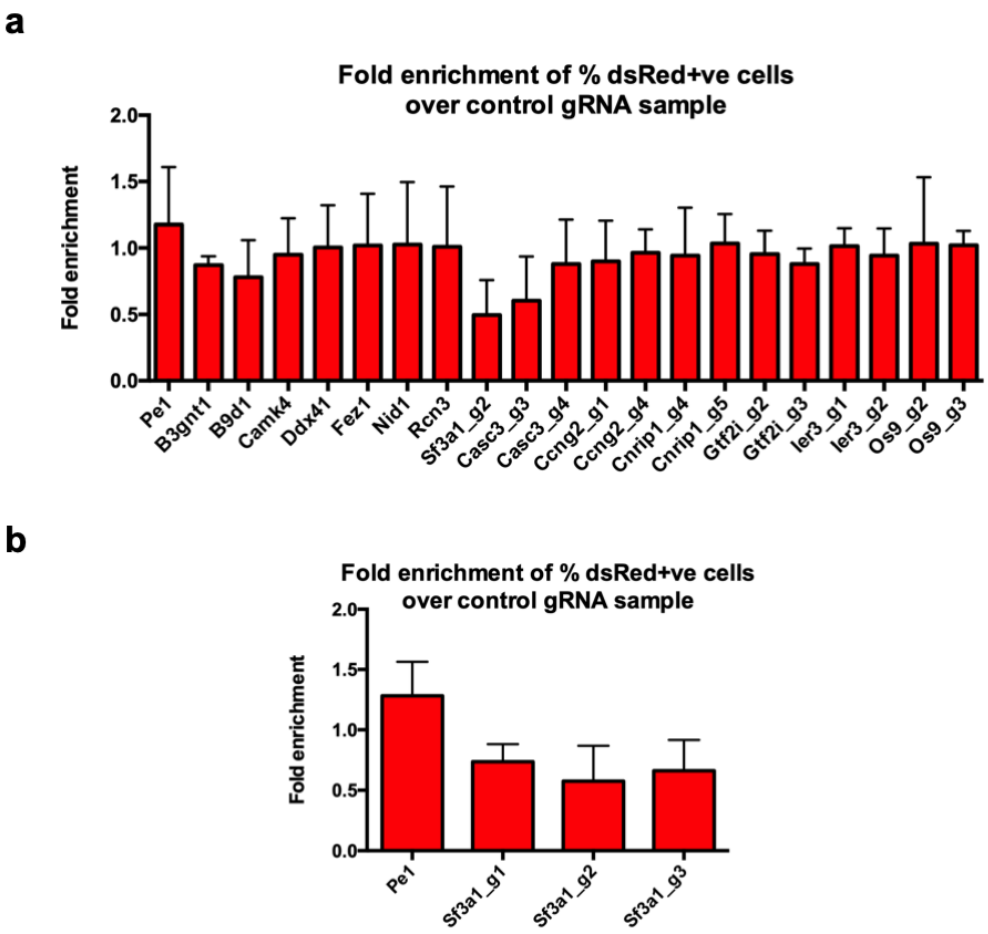


Fig. 4.10 – Individual testing of the selected gRNAs from the screen. (a) Fold enrichment over Zeomycin non-targeting gRNA of % iNs at day 15 upon transduction of *Pecam1* gRNA or gRNAs against genes listed in Fig.4.9 (n=2). **(b)** Fold enrichment over Zeomycin non-targeting gRNA of % iNs at day 15 upon transduction of *Pecam1* gRNA or *Sf3a1* gRNAs #1,2 and 3 (n=3). Error bars indicate standard deviation.

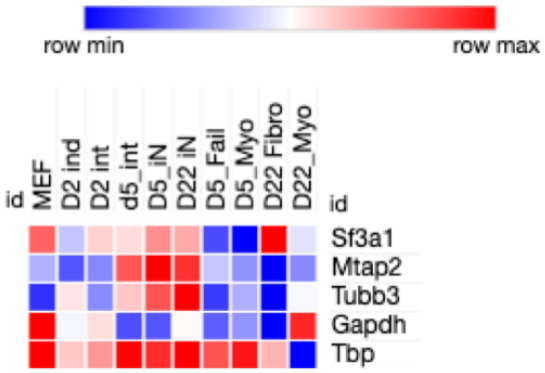
4.2.7 *Sf3a1* KO decreases iN conversion efficiency

Sf3a1 is a protein part of the spliceosome complex, specifically of the small nuclear ribonucleoprotein particle U2. It was first identified in 1995 as the mammalian homologue of the yeast essential splicing factor *Prp21p* and the Drosophila alternative splice regulator "suppressor-of-white-apricot" (Krämer et al., 1995). In human cells, it interacts with U1-SL4 to allow spliceosome assembly on pre-mRNA (Sharma et al., 2014) and it's required for HeLa cell survival, as its depletion leads to

severe defects in gene expression and consequent cell death (Tanackovic & Krämer, 2005). SNPs in its coding sequence have also been associated with several types of cancer, including pancreatic, colorectal (Chen *et al.*, 2015; Tian *et al.*, 2015).
It is known to be involved in modulating the strength of the innate immune response by regulating the production of alternative inhibitory splice forms of proteins involved in Toll-like receptor signalling in macrophages (De Arras & Alper, 2013; O'Connor *et al.*, 2015).
In neurons, it interacts with P2X6, a nuclear receptor protein suggested to recruit SF3A1 and impair its splicing abilities in aging mice (Díaz-Hernández *et al.*, 2015).
It has also been implicated to have a role in cell identity conversions: depletion of *Sf3a1* decreased the efficiency of reprogramming fibroblasts to cardiomyocytes, as well as decreasing the total number of cells (Zhou *et al.*, 2018).

In the MEF to iN conversion, it is expressed quite broadly in all cell types except myoblast-like cells at day 5 and 22 (Fig.4.11). Reported in Fig.4.11b are the normalised read counts for plasmid samples and neuron samples, together with log fold change and FDR. gRNA #2, the one that I originally selected based on FDR, is actually overrepresented in iNs by 2-fold (1 $\log_2$ fold change). Based on the gRNA's behaviour when tested individually, we can safely say that this is clearly an artefact of the screen conditions, and one example of how the screen noise is too high for even low FDRs to be reliable. The other two gRNAs' read counts show virtually no change. This is confirmation that the screen results are unreliable and overall the screening process did not work.

a



b

gRNA	Gene	Plasmid counts	iN counts	Plasmid mean	iN mean	LFC	FDR
Sf3a1_g1	Sf3a1	775.01/812.3/792.02	706.31/667.16/647.41	792.02	667.16	-0.24715	0.99797
Sf3a1_g2	Sf3a1	727.18/719.64/723.04	1493.7/1599.4/923.2	723.04	1493.7	1.0457	0.033859
Sf3a1_g3	Sf3a1	700.01/638.97/655.33	678.13/721.74/614.86	655.33	678.13	0.04927	0.99797

Fig. 4.11 – Sf3a1 expression and screen results. (a) *Sf3a1* expression pattern across the iN conversion (Treutlein *et al.*, 2016). Pan neuronal markers *Map2* and *Tubb3*, and housekeeping genes *Gapdh* and *Tbp* are shown as reference. **(b)** Normalised read counts from *Sf3a1* gRNAs #1,2 and 3 in the plasmid replicates and neuron replicates. LFC: Log fold change. FDR: False discovery rate.

I made use of the CRISPR/*Cas9* genome-wide screen performed by Dr. Kaemena in our lab (Kaemena, Beniazza et al., submitted for publication) to check the behaviour of *Sf3a1* gRNAs in the context of MEF to iPSC (induced pluripotent stem cell) reprogramming. Briefly, similarly to the screen discussed here, a genome-wide gRNA library was transduced into MEFs that were then reprogrammed into iPSCs. iPSCs were sorted at the end of reprogramming and the representation of gRNAs was verified through NGS. Fig.4.12a below shows the results of their screen. Compared to the plasmid library control, all 5 gRNAs against *Sf3a1* are severely depleted in

iPSCs. This could be due to *Sf3a1* being required for cell survival, or proliferation or being essential for the reprogramming process itself. Though the amount of information is limited, it serves as another piece of information that identifies *Sf3a1* as necessary for cell growth and possibly cell identity changes.

I then wanted to verify whether *Sf3a1* was having an effect on MEFs in the neuronal conversion context. I transduced *Cas9*-MEFs with either non-targeting gRNA Zeomycin or the three *Sf3a1* gRNAs individually, at MOI5, and cultured them in conversion medium for 15 days. I tracked the whole well cell numbers across this two-week period to check the effects of *Sf3a1* KO on these cells. Results are shown in Fig.4.12b. Though cell numbers overall decrease from day 6 to day 15, as cells stop proliferation and start dying off, *Sf3a1* KO cell numbers by day 6 are already more than 2-fold lower than control. This difference is maintained at day 10 and day 15. *Sf3a1* KO is clearly having an impact on proliferation, though this data is not enough to distinguish between slower cell proliferation in the first 6 days, or cell death.

a

sgRNA	Gene	Plasmid library	iPS 1	iPS 2	iPS 3
Sf3a1_g1	Sf3a1	389	67	19	69
Sf3a1_g2	Sf3a1	418	1	0	0
Sf3a1_g3	Sf3a1	16	0	0	0
Sf3a1_g4	Sf3a1	181	17	2	0
Sf3a1_g5	Sf3a1	165	46	177	17

b

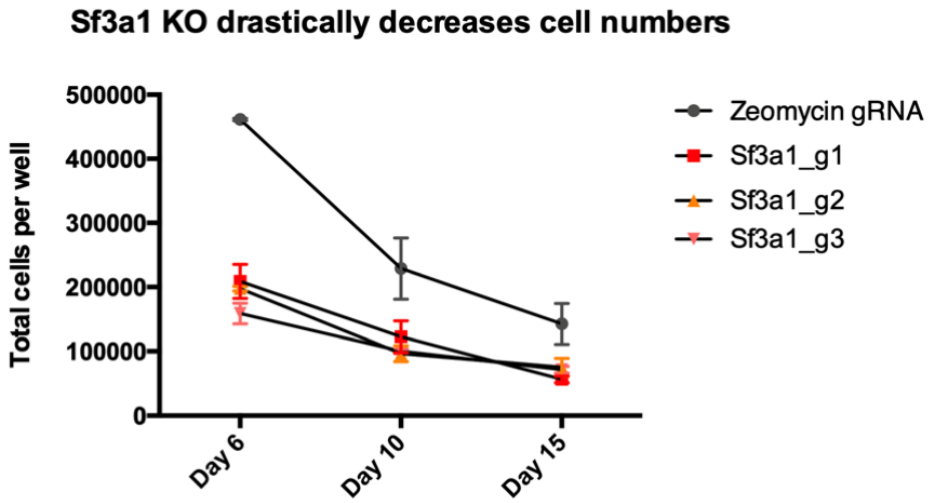


Fig. 4.12 – Sf3a1 KO decreases cell numbers. (a) Normalised read counts from the iPSC reprogramming CRISPR/*Cas9* screen performed in the lab (Kaemena, Beniazza et al., submitted for publication). All 5 *Sf3a1* gRNAs are shown. In yellow, read counts from the plasmid library, in green read counts from each of the 3 replicates of iPS. All 5 gRNAs show severe depletion in the iPS samples, indicating they are required for survival or proliferation of the cells. **(b)** *Cas9*-MEFs were transduced with gRNAs against Zeomycin or *Sf3a1* at MOI5 and then cultured in conversion medium for 15 days. Reported are whole well cell numbers for day 6, 10 and 15. *Sf3a1* KO results in a drastic reduction (more than 2-fold) in cell numbers from day 6 already that is maintained through day 15 (n=2). Error bars indicate standard deviation.

Sf3a1 seems to be having a double role in the iN conversion: KO of *Sf3a1* is decreasing total cell numbers, as well as decreasing the percentage of cells that manage to convert into neurons. Further investigation would be required to understand how these phenotypes are brought about. Considering its role in aging neurons, it would

be interesting to investigate its associations with senescence pathways. Moreover, considering its role in modulating alternative splicing in macrophages, it would be interesting to test whether overexpression of *Sf3a1* in the MEF to iN conversion might influence the conversion efficiency through the stabilisation of certain mRNA isoforms. A more thorough discussion of the relevance of this gene in iNs is described in chapter 6.

4.3 Discussion

4.3.1 Screening tools and conditions could not be verified and likely contributed to the failure of the screen

Figures 4.8 and 4.11 showing read counts for *Ascl1*, *p53* and *Sf3a1* gRNAs within the screen results clearly indicate that the screen conditions were not suitable for detection of these phenotypes. There are a few technical issues that might have contributed to this.

Firstly, the lack of a proper set-up for a pilot screen made it difficult to optimise the screening conditions. Specifically, I was not able to find an appropriate control gene that, upon knock-out, would give a known and predictable phenotype that I could try to pick-up on using the expected screen coverage. A recent publication showed how *Kmt2b* is required for efficient conversion of MEFs into iNs and how *Kmt2b* knock-out MEFs show deficiencies in their conversion ability (Barbagiovanni *et al.*, 2018). I've verified this phenotype in my system (discussed in the next chapter). For the purpose of a future screen, it could be used as control gene to verify the coverage requirements for detection of a depletion phenotype.

Secondly, it's worth mentioning that the MEFs used in this study express an *SpCas9* sequence that contains a single nuclear localization signal (NLS). It has been shown how a double NLS improves the protein nuclear localization in mammalian cells and improves its editing efficiency (Cong *et al.*, 2013), so this factor could have made the system less efficient than other reported screens and made it harder to pick-up phenotypes from sub-optimal conditions like poor gRNA efficiency. Nonetheless, I've

shown how this *Cas9* can induce efficient KO when cells are transduced with gRNA at MOI5, as shown by *Ascl1* KO in Fig.4.2a and by *Sf3a1* KO experiments in figures 4.10-12. I've also shown how this *Cas9* can induce efficient KO when provided with a single gRNA copy lentivirally (Fig.4.2b), suggesting that lentiviral gRNA expression and gRNA efficiency can compensate for the absence of a second NLS.

Another consideration to keep in mind is gRNA competition. As mentioned before, each gRNA within the screen context will on average be present as a single expression construct within the transduced cells, and will have to compete with two other gRNA species also integrated as single constructs for *Cas9* loading. I have not addressed whether in this competition environment *Cas9* expression levels become a limiting factor in its editing efficiency, and whether this is made worse by its sub-optimal nuclear localization.

Moreover, it has been shown how *Cas9* targeting of heterochromatin regions is less efficient (Knight *et al.*, 2015; Singh *et al.*, 2015). Considering the library's bias towards neuronal genes that are mostly in a repressed closed state in MEFs, as assessed by ATACseq and FAIREseq (Wapinski *et al.*, 2013; Treutlein *et al.*, 2016; Wapinski *et al.*, 2017), it is possible that *Cas9* editing at those loci was less efficient. Many of the loci would have become open at 48h and then at day 5 post-induction of the TFs (Wapinski *et al.*, 2017), so they could have been potentially targeted then. Nonetheless, *Cas9* expression level changes during the reprogramming process have not been addressed. Though the *Rosa26* locus should be constitutively expressed, the chromatin conformational changes associated with the reprogramming process might lead to temporary or permanent silencing. There is also the possibility that, even If a locus became accessible at day 2 or day 5 for *Cas9* to edit, the cell would have already made its decision to convert into a neuron, and/or the protein of the gene knocked-out would still be around for long enough to allow conversion.

Overall, these considerations on the screening tools and starting conditions potentially explain the low power of the system to pick up verified phenotypes like *Sf3a1*.

4.3.2 Selective pressure improves screen data analysis

Cell proliferation plays an important role in the enrichment or depletion of gRNAs within the populations. Previous CRISPR/*Cas9* KO screening have used ESCs or cancer lines that grow indefinitely, and screened for phenotypes looking at cell fitness or resistance to a specific drug (Tzelepis *et al.*, 2016; Koike-Yusa *et al.*, 2014; Shalem *et al.*, 2014; Wang *et al.*, 2014). In this context, a gRNA targeting an essential gene will become depleted quite efficiently, as the cells harbouring neutral gRNAs (or gRNAs enhancing cell growth) will be propagated by repeated cell divisions, and thus the depletion will be easier to distinguish from background. Moreover, these type of cell lines can specifically be used to artificially boost gRNA depletion or enrichment by allowing the cells to grow for extended periods of time to give milder phenotypes enough cell doublings to become statistically relevant. One study kept proliferating cells in culture for 3 weeks and performed serial passaging to induce a strong selective pressure on gRNA representation (Doench *et al.*, 2016). In contrast, the experimental setup of the iN conversion is extremely limiting in this regard. Firstly, because the target cell population, the neurons, is post-mitotic, and stops dividing between day 3 and day 5 when the major wave of cell fate change is occurring. This means that a gRNA targeting a roadblock gene will be overrepresented in the iN population, but won't be expanded by the cell growth as it happened in the screen papers reported. Similarly, a gRNA targeting an essential gene will be depleted, but because the remaining of the cells won't out-grow the cells harbouring it, its depletion will struggle to become statistically relevant. This becomes more evident when we consider a second aspect, i.e. that overall all of the cells stop proliferating around 48h post-induction of the TFs. Even without the transduced BAM conditions, MEFs *in vitro* senesce quickly and arrest proliferation through contact inhibition mechanisms. The neuronal medium moreover lacks serum, another factor limiting MEF growth.

A clear example of the limitations of this system comes from the results of the *p53* gRNAs read counts within the non-neuronal population as compared to the library read counts. These gRNAs have been validated as functional and do induce significant

proliferation within previous screens in the lab (Kaemena, Beniazza *et al*., submitted) and in this thesis as well, although under optimal conditions such as MOI5 and no competing gRNAs (Fig.3.2). Nonetheless, in these experimental conditions, *p53* gRNAs are not enriched even within the population of cells that has the potential to proliferate. This indicates that either *p53* KO is not enough to overcome the non-proliferating conditions that the cells are put in (i.e. BAM overexpression might be blocking proliferation more than *p53* KO could enhance), or that the KO is inefficiently happening, or that there are too few cells that receive the gRNA and manage to proliferate, but the coverage of the screen is not enough to pick up on their growth.
Considering this set-up, and having identified potential controls to be used in pilot-settings, a smaller scale screening strategy would allow for higher gRNA coverage and increase the chances of identifying significantly depleted or enriched genes. For instance, using the same number of cells, the gRNA number could be reduced to 2000 (1/3rd of this library), to increase the gRNA coverage in iNs from 450 to 1350. It would also allow the use of a library transduction at MOI=0.3, so that each cell would only have 1 gRNA species. This would limit gRNA competition for *Cas9*, possibly increasing its editing efficiency, and would limit the potential background noise coming from neutral gRNAs being carried-over together with functional gRNAs. To clarify this with an example: the gRNAs that were transduced in the same cells as gRNAs targeting *Sf3a1* would have gotten depleted artificially as cells died off. Because these gRNAs likely were different in each cell, this process will have increased depletion background noise.

4.3.3 *Sf3a1*'s role in cell identity conversions and alternative splicing

As reported by the analysis of *Sf3a1* KO, this gene seems to be important for both cell fitness, since it decreases total cell numbers, and iN conversion efficiency, since it decreases the percentage of iNs obtained. This role doesn't seem to be specific for the MEF to iN system, as it has been already reported to be required for survival of human HeLa cells (Tanackovic & Krämer, 2005), survival of MEFs and induced

cardiomyocytes, as well as the efficiency of surviving MEFs to be converted into cardiomyocytes (Zhou *et al.*, 2018). Fig. 4.12 highlights its potential role in iPSC reprogramming, although it seems to be essential for mESC fitness in general, as revealed by the depletion of gRNAs targeting *Sf3a1* in mESC fitness CRISPR-*Cas9* loss-of-function screen performed by the Yusa lab (Tzelepis *et al.*, 2016).

It also seems to be required for fitness in five different immortalised and cancer human cell lines (Hart *et al.*, 2015).

In the context of cell identity conversions, alternative RNA splicing has been associated with regulating pluripotency and differentiation. The MBNL proteins have been identified as negative regulators of pluripotency splicing profiles, and their knock-down leads to an upregulation of ESC-like alternative splicing events (Han *et al.*, 2013). MBNL proteins regulate an alternative splicing event that generates an ES-specific form of FOXP1. This isoform upregulates pluripotency-related genes and suppresses differentiation genes, and its knock-down drastically reduces the efficiency of MEF to iPSC conversion (Gabut *et al.*, 2011). Alternative splicing has further been characterised in IPSCs: converting MEFs switch from a somatic to a pluripotent splicing profile, and the knock-down of two RNA-binding proteins U2AF1 (also part of the U2 splicing complex) and SRSF3 decreases IPSC reprogramming efficiency (Ohta *et al.*, 2013).

Sf3a1 has not been extensively characterised, but one relevant study has highlighted its role in innate immune response in mice (O'Connor *et al.*, 2015). The authors showed how *Sf3a1* is required for robust immune response by inhibiting the production of alternatively spliced signalling adaptor *MyD88*, that would normally work within a negative feedback loop to dampen the immune activation. They also showed how *Sf3a1* knock down in this system does not impact cell viability, but weakens the innate immune response of murine macrophages. This suggests that *Sf3a1* works, at least within these cells, by preventing alternative splicing of negative regulators of immunity.

Based on this information, it would be very interesting to characterise alternative splicing in the conversion of MEFs to induced neurons, with particular attention to

Sf3a1. Firstly, it would be worth investigating the difference between knock-out and knock-down, to check whether knock-down might allow enough SF3A1 protein levels to maintain cell proliferation, but impact conversion.

Secondly, overexpressing *Sf3a1* during the iN conversion could potentially stabilise certain protein isoforms required for neuronal identity and enhance the conversion efficiency. It would also be interesting to characterise the pre-mRNA species bound by *Sf3a1* in MEFs and iNs using technologies like Cross-linking immunoprecipitation (CLIP) to identify alternatively spliced genes that would be otherwise be harder to distinguish using traditional RNAseq.

4.3.4 *Sf3a1* – clues to aging neurons

SF3A1 has been identified as a binding partner of the purinergic receptor subunit P2X6 in mouse hippocampal neurons (Díaz-Hernández *et al.*, 2015). As P2X6 localises to the nucleus, it recruits and partly sequesters *SF3A1*, impairing splicing events. As the expression and nuclear localisation of both genes is increased in aging mice, and P2X6 overexpression leads to a decrease in splicing events, it has been suggested that P2X6 might be partly responsible for defective splicing in aging neurons. Pre-mRNA splicing defects have long been associated in fact with aging and senescence (reviewed in Meshorer and Soreq, 2002).

In the context of induced neurons, it would be interesting to investigate whether P2X6 expression levels are different between MEFs and adult mouse fibroblasts, or between human fibroblast of embryonic or neonatal origin versus adult and possibly aged patients. We can speculate how excessive levels of this protein might be impairing iN reprogramming efficiency by sequestering *SF3A1*. Conversely, P2X6 knock-out might lead to an increased availability of *SF3A1* that could potentially mimic its overexpression – consequences of which have been considered in the previous section.

Overall, the role of *Sf3a1* in neurons has not been extensively addressed, hence it could hold potential as a starting point to better understand the relationship between mRNA splicing and neuronal fate.

4.4 Summary

This chapter discusses CRISPR/*Cas9* screen itself, the analysis of the sequencing results and the validation of *Sf3a1* as an essential gene for iN conversion.

The first part focuses on pilot-testing the conditions for the screen, using *Ascl1* KO and *Tubb3* KO phenotypes to try and predict in small scale the behaviour of gRNA representation in iNs after the conversion. Unfortunately, both *Ascl1* and *Tubb3* proved to be unsuitable controls to knock-out to perform a pilot testing.

The second part focuses on the experimental screen conditions and the analysis of the sequencing results. MAGeCK, the algorithm used to rank gRNA representation and identify gene hits, reported that no gene was significantly (FDR <0.05) enriched or depleted when comparing neurons with fibroblasts, neurons with plasmid and fibroblasts with plasmid. This is probably due to the intrinsic high variability between neuronal and, to a smaller extent, non-neuronal samples, as well as the coverage not high enough to pick up on minor phenotypes due to the absence of selective pressure against neutral gRNAs or biological amplification of gRNAs that produce a phenotype. A few gRNAs were selected based on their individual FDR and individually tested at MOI5. One gRNA against *sf3a1* resulted in 50% decrease in neuronal conversion efficiency when tested individually. Two more gRNAs against *Sf3a1* confirmed this phenotype.

The third part focuses on validating and contextualising the *Sf3a1* KO phenotype. *Sf3a1* is part of the pre-mRNA spliceosome complex, and it is known to be involved in cell survival. It has been shown to have a role in fibroblast to cardiomyocyte

conversion. In the context of iNs, it is expressed in neurons and fibroblast-like cells and, upon KO, reduces the conversion efficiency to neurons by about 50%. It also significantly decreases the cell numbers of MEFs alone, though further characterisation would need to address whether cells are dying or proliferating at a slower pace.

The next and final results chapter takes a parallel approach to screening and focuses on testing individually selected genes from literature one by one.

Chapter 5 Investigating individual candidate genes as potential regulators of the iN conversion

5.1 Introduction

In addition to the screening, I aimed to test the KO of several candidate genes in order to identify important regulators of the iN conversion. Considering the differences between neuronal differentiation and iN conversion (for instance that during MEF to iN conversion cells do not go through a neural stem cell state (Treutlein *et al.*, 2016)), I decided to select candidate genes exclusively based on iNs papers, reasoning that those would be more likely to result in relevant hits. I used two different approaches to select the candidate genes: first, I looked for genes reported to be targets of *Ascl1* or to have a role as facilitators of the conversion. *Ascl1* is the pioneer master TF of the conversion and is able to reprogram fibroblasts to neurons alone, although at a lower efficiency and yielding immature neurons (Wapinski *et al.*, 2013; Vierbuchen *et al.*, 2010). I hypothesised that knocking out its targets could potentially impair *Ascl1* function and hinder the conversion.

Second, I selected candidate genes from the 1203 genes included in the library based on their expression within the first 5 days of the conversion. I selected the genes that were consistently differentially expressed between day 2 induced cells (cells that had high *Ascl1* activation levels, see Table 3.1 for more detailed explanation) and MEFs, day 5 neurons and day 2 induced cells, and day 5 neurons vs non neuronal cells. This served the secondary purpose of identifying potential internal library controls: if the KO of any of these genes resulted in a clear phenotype (increase or decrease in neuronal conversion efficiency), I could use this information to examine the read counts of the gRNAs targeting that gene within the screening sequencing results and assess the screen's ability to pick up known phenotypes.

In total, I tested 20 individual genes. The knock-out of one of these, *Stxbp1*, resulted in a reduction of neuronal conversion efficiency, which lead me to investigate *Stxbp1*'s role in the iN system.

5.1.1 Aims of this chapter

The first half of this chapter aims to investigate the possible role of 20 individual genes in the iN conversion. The second half focuses on characterizing the function of *Stxbp1* – the knock-out of which decreases iN conversion efficiency by about 50%.

5.1.2 Notes

1) The genes analysed in this chapter were labelled for convenience with numbers ranging from 1 to 19 – and they are reported with this labelling in Fig.5.3. Table 5.1 indicates the correspondence between genes and numbers.

Gene number	**Gene name**	**Molecular function**	**Role in the conversion**
1	*Dl11*	Transmembrane ligand protein	*Ascl1* target based on ChIPseq
2	*Hes6*	Transcription factor	*Ascl1* target based on ChIPseq
3	*Lmo2*	Transcription factor	Increases efficiency of *Myt1l* + *Zfp238* reprogramming
4	*Mfng*	Glycosyltransferase	*Ascl1* target based on ChIPseq
5	*NeuroD4*	Transcription factor	*Ascl1* target based on ChIPseq
6	*Olig2*	Transcription factor	Increases efficiency of *Ascl1*-only reprogramming
7	*Brn4 (Pou3f4)*	Transcription factor	Increases efficiency of *Ascl1*-only reprogramming
8	*Rfx1*	Transcription factor	Increases efficiency of *Myt1l* + *Zfp238* reprogramming

9	*TcfL5*	Transcription factor	Increases efficiency of *Myt1l* + *Zfp238* reprogramming
10	*Zfp238 (Zbtb18)*	Transcription factor	*Ascl1* target, makes iNs when expressed with *Myt1l*
11	*Zic1*	Transcription factor	Increases efficiency of *Ascl1*-only reprogramming
12	*2700081O15Rik*	Uncharacterised	Upregulated
13	*Anxa3*	Inhibitor of phospholipases	Downregulated
14	*Cnn2*	Filament-associated protein	Downregulated
15	*Dcaf5*	E3 ubiquitin-protein ligase complex receptor	Upregulated
16	*Fam171b*	Uncharacterised	Upregulated
17	*Gja1*	Gap junction protein	Downregulated
18	*Stxbp1*	Syntaxin binding protein involved in vesicle docking	Upregulated
19	*Vmac*	Vimentin-associated coiled coil protein	Upregulated

Table 5.1 – Gene number codes and descriptions. Information on the genes' role in the conversion is taken from Vierbuchen et al., 2010 and Wapinski et al., 2013.

2) As for chapter 4, results of gene KO in conversion experiments assessed by flow cytometry are reported as the fold-enrichment of the percentages of *hSyn1*-dsRed+ve cells over samples transduced with the Zeomycin control gRNA. The second control, reported together with the rest of the samples in the graph, is the intronic gRNA for *Pecam1*.

5.2 Results

5.2.1 Selecting genes to individually test based on their function and gene expression

The first 11 genes I selected were chosen based on their implication in the iN conversion from previously published work (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013, 2017). Fig5.1 summarizes their expression patterns and levels relative to the neuronal markers *Tubb3* and *Map2*, and the housekeeping genes *Gapdh* and *Tbp*.

Zic1, *Brn4 (Pou3f4)* and *Olig2* were selected because, when overexpressed, they increase the conversion efficiency of *Ascl1*-only reprogramming (highlighted in green hues in Fig5.1b).

Dl11, *Hes6*, *Mfng* and *NeuroD4* were selected because they are targets of *Ascl1* based on ChIP-seq analysis and upregulated during the first phases of iN conversion (highlighted in red hues in Fig5.1b).

Zfp238 is a direct target of *Ascl1* and is able to reprogram MEFs to iNs in the absence of *Ascl1* when combined with *Myt1l*. Its KO leads to a significant decrease in *Ascl1*-only conversion efficiency. Adding *Lmo2*, *Rfx1* or *Tclf5* to the *Zfp238*+*Myt1l* reprogramming cocktail greatly increases their conversion efficiency (highlighted in blue hues in Fig5.1b).

These 11 genes could potentially be facilitators of the conversion and their KO might lead to a decrease in neuronal conversion efficiency.

As shown in the heatmap, the majority of the genes' expression increases in neuronal cells as the conversion progresses, with the exception of *Zic1* (showing detectable expression only in day 22 fibroblasts (based on this dataset) and *NeuroD4* and *Olig2,* that show consistently low expression across the conversion. The raw expression levels, as shown in Fig.5.1b, are quite variable. Only *Hes6*, *Mfng* and *Zfp238* show significant levels at day 2, day 5 and day 22 cells above the housekeeper *Tbp* and similar to *Tubb3* pattern.

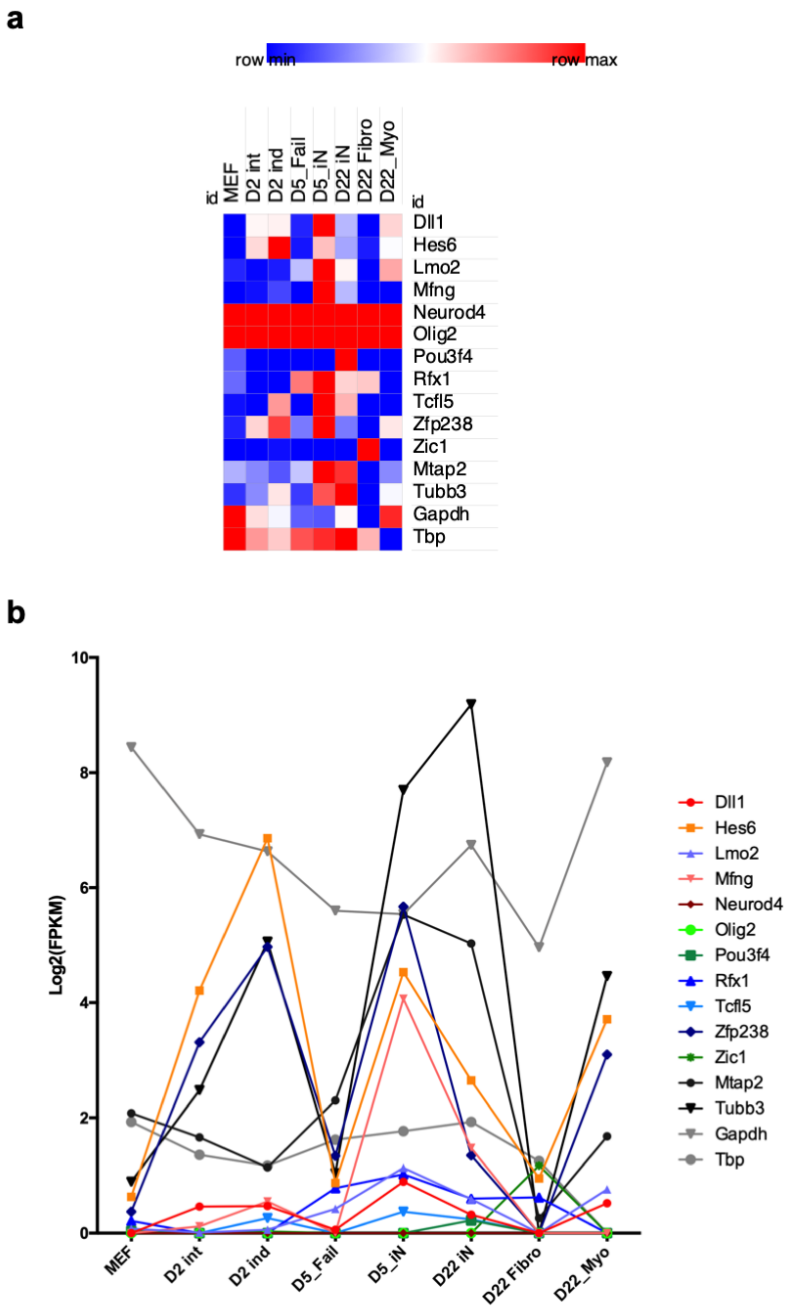


Fig. 5.1 – Expression patterns and levels of the first 11 candidate genes. (a) Heatmap showing the expression patterns of the 11 genes across the different cell populations analysed by Treutlein and colleagues (Treutlein *et al.*, 2016). The neuronal markers *Tubb3* and *Map2*, and the housekeeping genes *Gapdh* and *Tbp* are included as reference. **(b)** Expression levels – from the same dataset – of the 11 genes. Highlighted in green hues are conversion facilitators *Zic1*, *Brn4* and *Olig2*. Highlighted in red hues are *Ascl1* targets *Dl11*, *Hes6*, *Mfng* and *NeuroD4.* Highlighted in blue hues are *Zfp238* and *Lmo2*, *Rfx1* and *TcfL5*, which increase conversion efficiency when co-expressed with *Zfp238* and *Myt1l*.

Further 8 genes were selected from the screening library itself, based on expression levels. This meant that I could also potentially use any validated phenotype as an internal control for the screen. I focused on the initial steps of conversion – up to day 5, reasoning that a gene likely to have a big impact on the conversion would show strong upregulation or downregulation in the early phases of the establishment of iNs. I used the comparisons between day 2 induced cells vs MEFs, day 5 iNs vs day 2 induced cells, and day 5 iNs vs day 5 *Tau*-eGFP-ve cells (more details about this categorisation of cell populations are found in Table 3.1).

2700081O15Rik, *Dcaf5*, *Fam171b*, *Stxbp1* and *Vmac* were chosen because they are consistently upregulated (>2 fold-change) in all three comparisons.

Anxa3, *Cnn2* and *Gja1* were chosen because they are consistently downregulated (<0.3 fold-change) in all three comparisons.

Expression patterns and levels of these 8 genes are shown in Fig.5.2a,c and Fig.5.2b,d respectively.

Of the upregulated genes, only *Fam171b* and *Stxbp1* maintain high expression in day 22 iNs, though *Fam171b* is also expressed in day 22 myoblasts.

All three downregulated genes show very little expression in day 22 iNs and show high expression in day 22 fibroblasts instead.

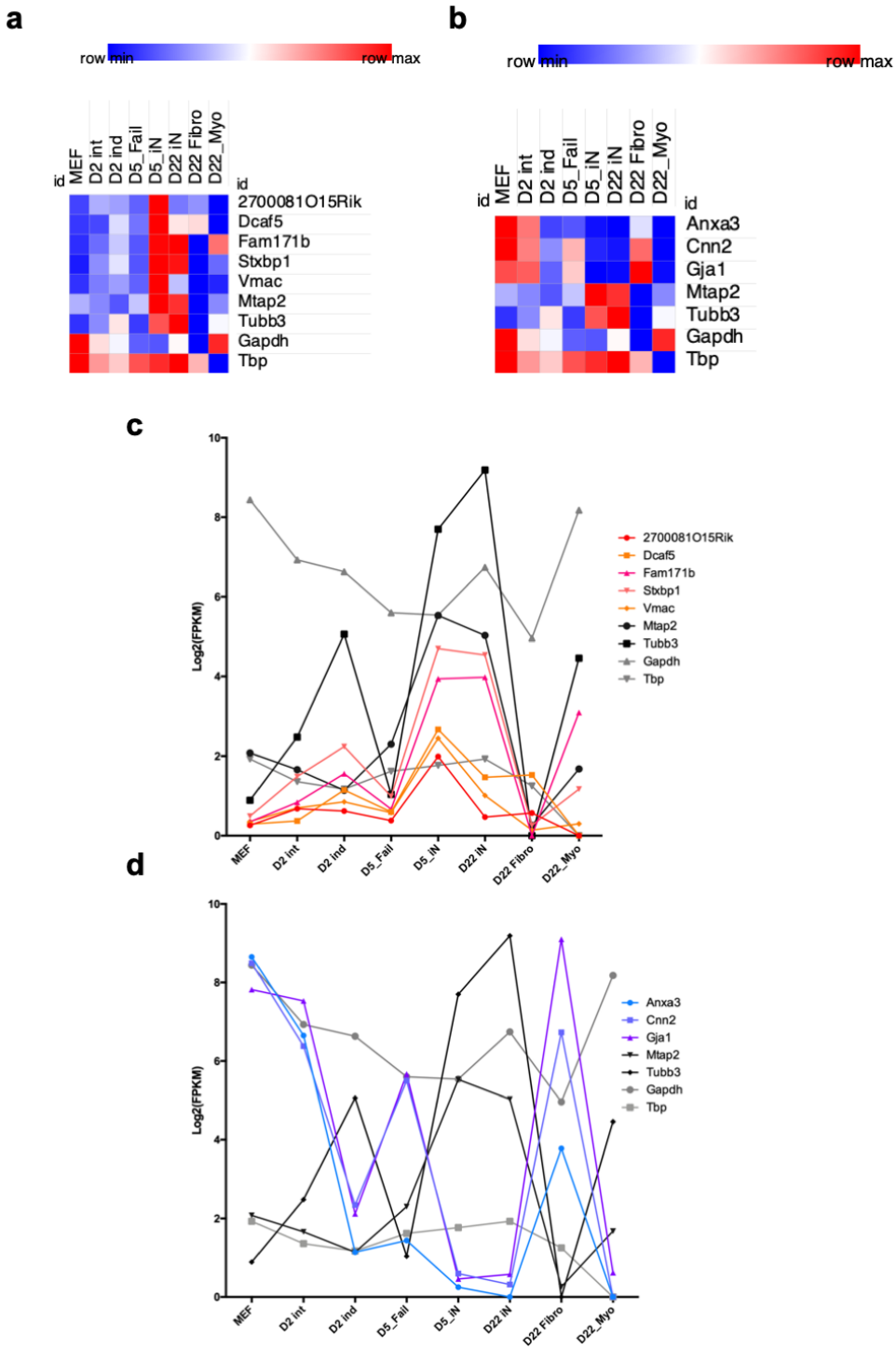


Fig. 5.2 – Expression patterns and levels of the remaining 8 candidate genes. (a, b) Heatmap showing the expression patterns of the 8 genes across the different cell populations analysed by Treutlein and colleagues (Treutlein *et al.*, 2016). The neuronal markers *Tubb3* and *Map2*, and the housekeeping genes *Gapdh* and *Tbp* are included as reference. **(c)** Expression levels – from the same dataset – of the 5 upregulated genes. **(d)** Expression levels – from the same dataset – of the 3 downregulated genes.

I additionally tested one more gene: *Kmt2b*. This gene was shown in October 2018 to be required for the conversion of MEFs into iNs (Barbagiovanni *et al.*, 2018). In the paper, they use genetically engineered MEFs from a mouse line where exon 2 of *Kmt2b* is flanked by *LoxP* sites and KO is achieved by expression of *Cre* recombinase upon tamoxifen-driven induction. They use PSA-NCAM to quantify iNs by flow cytometry, and *Tubb3* to quantify iNs by immunofluorescence, and they report a 33% and 66% reduction in iNs upon *Kmt2b* KO as assessed by *Tubb3+ve* or PSA-NCAM +ve cells respectively.
By testing the effects of *Kmt2b* KO in the iN conversion myself I wanted to validate the efficiency of my system in reproducing published phenotypes while using different genetic tools for KO and iN selection.

3 gRNAs per each of these 20 genes were chosen from the same genome-wide library used in Chapters 3 and 4 (Tzelepis *et al.*, 2016) and cloned within the pLV-U6-gRNA backbone for lentiviral delivery.

5.2.2 *Kmt2b* KO reduces the percentage of iNs generated

Each gRNA was tested individually in a conversion experiment by transducing converting MEFs at day -1 with lentiviral particles at MOI5 together with the BAM factors. At day 3 converting cells were transduced with *hSyn1-dsRed* to allow quantification of iNs via flow cytometry.
Fig.5.3 summarises the results of testing these 20 genes. Panel **(a)** shows the results of *Kmt2b* KO with three different gRNAs. The KO induces a mild reduction (about 30% on average) in the % of dsRED+ve cells by day 15 of the conversion for all three gRNAs tested. This is confirmation that the system I developed using CRISPR/*Cas9* to induce KO and *hSyn1*-dsRed to mark iNs can reproduce published phenotypes.

5.2.3 *Stxbp1* shows a 50% reduction in iN conversion

Results of knocking-out the other 19 genes are shown in Fig.5.3b-c. With the exception of gene #18, all the other genes tested fail to show any significant change in dsRED % upon KO, with most gRNAs yielding a fold-enrichment over the control sample (Zeomycin gRNA) between 1 and 1.2. It has to be mentioned that, not having verified each gRNA's ability to induce indels and lead to KO, it is possible that some of these genes failed to be deleted, and thus these results could be false negatives. Thus we can't claim that KO of any of these genes does not affect the conversion.
It is surprising that the KO of *Zfp238* did not decrease the percentage of dsRED+ve iNs. 3 major factors are different between the work by Wapinski and colleagues (Wapinski *et al.*, 2017) and the experimental set-up here: the use of *Zfp238* KO MEFS vs gRNAs, the use of *Ascl1*-only reprogramming vs BAM reprogramming cocktail, and the quantification of cells using TUBB3/MAP2 staining vs *hSyn1-dsRed* flow cytometry. Assuming gRNAs are correctly inducing KO, it's possible that *Zfp238* is not required in the presence of *Brn2* and *Myt1l*. It's also possible *hSyn1-dsRed* is marking a broader subset of iNs that, if analysed by staining, would show failure to upregulate *Tubb3* and *Map2* in the absence of *Zfp238*.

The KO of *Stxbp1* (gene 18 based on the labelling in Table 5.1) instead shows a remarkable phenotype, resulting in an average 50% decrease in dsRED +ve cells at day 15.

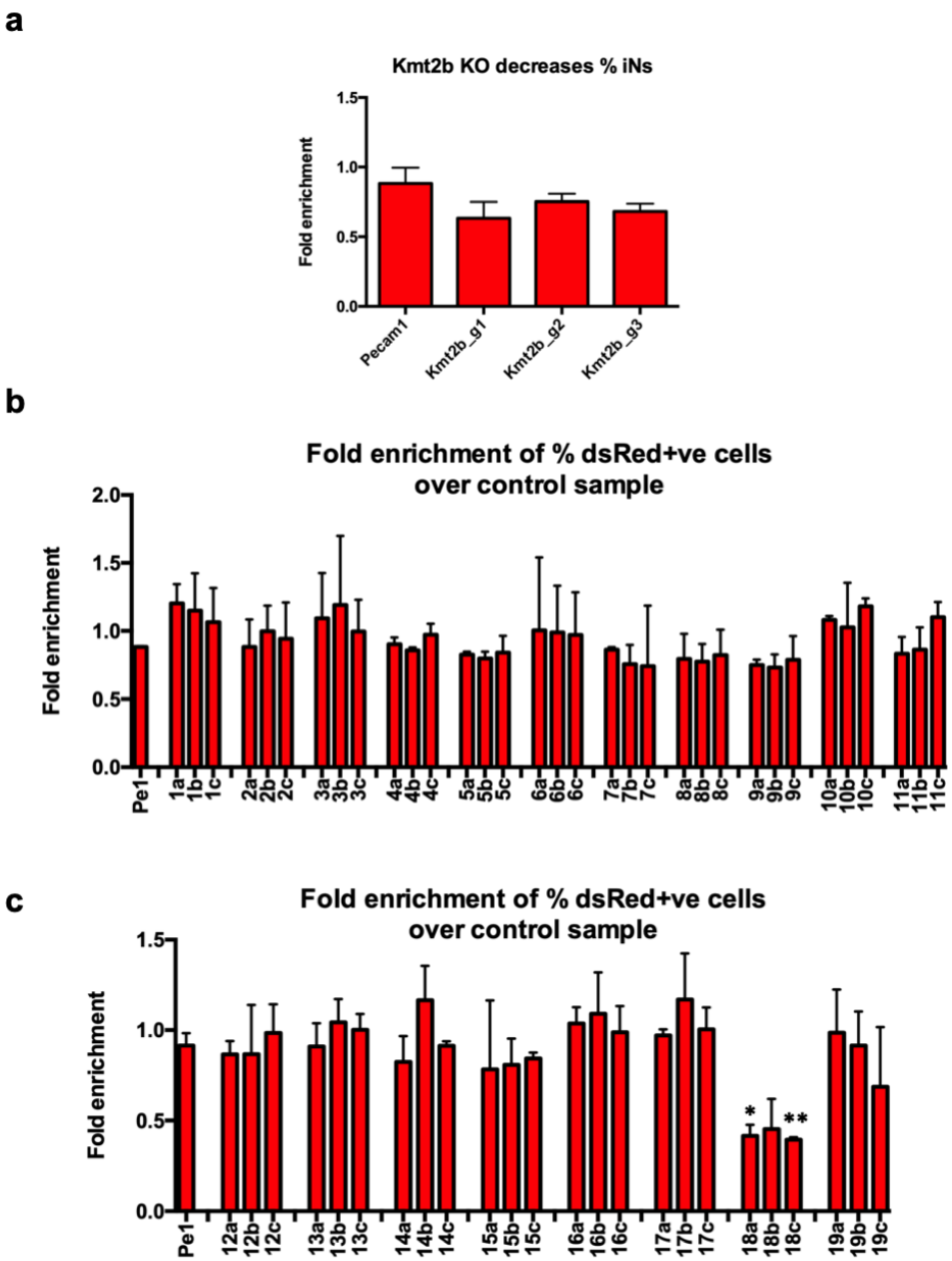


Fig. 5.3 – Kmt2b and 19 candidate genes KO results assessed by hSyn1-dsRed expression via flow cytometry. (a) Fold enrichment over Zeomycin non-targeting gRNA of % iNs at day 15 upon transduction of *Pecam1* gRNA or *Kmt2b* gRNAs (n=3). **(b)** Fold enrichment over Zeomycin non-targeting gRNA of % iNs at day 15 upon transduction of *Pecam1* gRNA or gRNAs against genes 1-11 in Table 5.1 (n=2). **(c)** Fold enrichment over Zeomycin non-targeting gRNA of % iNs at day 15 upon transduction of *Pecam1* gRNA or gRNAs against genes 12-19 in Table 5.1 (n=2). KO of Gene #18 – *Stxbp1* – shows a 50% decrease in the % of iNs. Error bars indicate standard deviation. ***p<0.005, **p<0.01, *p<0.05 based on an unpaired two-tailed Student's t test.

Fig.5.4a shows the validation of each of *Stxbp1* 3 gRNAs in 3 biological replicates. Each guide shows a consistent effect – within itself and among the other 2 two guides. Considering all three guides are giving the same phenotype, it is safe to assume the effect is coming from *Stxbp1* KO and not off-target effects of the gRNAs, which would be unlikely to result in the same phenotype from 3 different guides.

Nonetheless, to confirm *Stxbp1* KO at DNA sequence level, I selected gRNA #3, the one giving the strongest phenotype, and assessed the presence of indels using TIDE (Brinkman *et al.*, 2014). This software allows the deconvolution of *Cas9*-edited pooled sequences and calculates the percentage of indels generated by a given gRNA. I transduced MEFs with gRNA #3 at MOI5, then extracted the genome after 72h (when *Cas9* editing will have fully occurred (Tzelepis *et al.*, 2016; Koike-Yusa *et al.*, 2014)) and PCR-amplified a 500bp region containing the gRNA target sequence from edited MEFs (*wt* sequence amplified from unedited MEFs was used as a reference in the TIDE algorithm). This region was then sent for Sanger sequencing and the resulting chromatograms were inputted into the TIDE web tool (https://tide.deskgen.com/ - Accessed on 20/10/18).

The results are shown in Fig.5.4b. gRNA #3 is correctly editing exon 1 of *Stxbp1* with an indel efficiency of 69.4%.

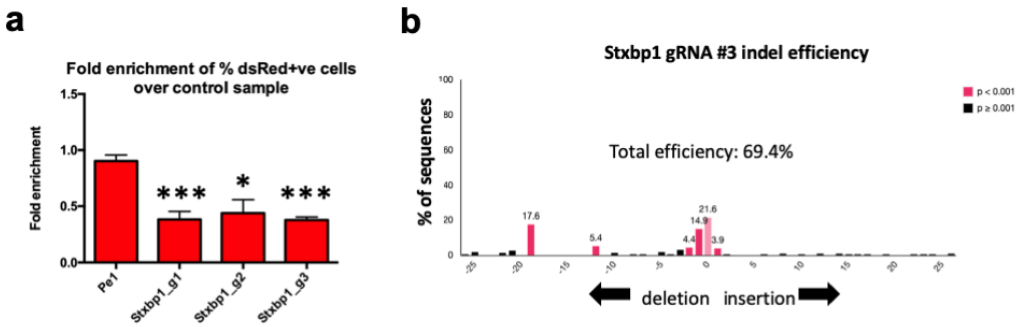


Fig. 5.4 – *Stxbp1* KO results. (a) % of dsRED+ve cells at day 15 of the conversion. Shown is the fold enrichment over a control (Zeomycin) non-targeting gRNA. All 3 gRNAs tested for *Stxbp1* show a circa 50% reduction in iNs compared to the *Pecam1* (Pe1) control gRNA (n=3). Error bars indicate standard deviation. ***p<0.005, **p<0.01, *p<0.05 based on an unpaired

two-tailed Student's t test. **(b)** Indel efficiency of *Stxbp1* gRNA #3 was assessed by TIDE. X axis indicates the numbers of deleted or inserted nucleotides.

5.2.4 *Stxbp1* screen results

Although the phenotype of *Stxbp1* KO was confirmed for 3 independent gRNAs, and gRNA indel efficiency was confirmed, the screen set-up failed to pick-up *Stxbp1* as having an effect on iN conversion. Table 5.2 below shows the read counts of all 3 individually tested *Stxbp1* gRNAs in the plasmid and iN samples in the screen. The read counts show high variability among replicates in the neuron samples, and the resulting means actually show a mild increase in representation of the gRNAs in neuronal samples compared to plasmid. This is further proof that the screen setup was inadequate for detection of even a 50% depletion phenotype, and it's probably due to the low coverage and the high variability of the system.

gRNA	Plasmid counts	iN counts	Plasmid mean	iN mean	LFC	FDR
*Stxbp1*_1	1191.3/1210.9/1186.7	1426.7/2176.4/1567	1191.3	1567	0.39515	0.99797
*Stxbp1*_2	130.44/142.97/153.29	177.9/218.3/124.78	142.97	177.9	0.31338	0.99797
*Stxbp1*_3	807.62/865.01/838.01	957.31/1439/906.92	838.01	957.31	0.19181	0.99797

Table 5.2 – gRNA counts from screen results. gRNA read counts from three replicates of plasmid and iNs samples are shown for all 3 gRNAs against Stxp1 that were individually tested. LFC: Log fold change; FDR: False discovery rate. LFC is calculated from iN mean over Plasmid mean. The LFC actually shows a minor overrepresentation of the gRNAs in iNs, but the FDR for all 3 of them is >0.95.

5.2.5 *Stxbp1* regulates neurotransmitter released and its knock-out leads to neuronal death

Stxbp1, also known as *Munc18-1*, is a syntaxin-binding protein required for neurotransmitter release by neurons. It was first discovered as a protein binding the N-terminus of *Syntaxin1A* in pre-synaptic vesicle fusion protein complexes. (Hata, Slaughter & Südhof, 1993). It is expressed almost exclusively in the brain in rodents and humans (Garcia *et al.*, 2003; Kalidas *et al.*, 2000), and it is localised preferably within axons in membranes proximity, thanks to its association with *Syntaxin1* and other binding partners within the vesicle fusion complexes (Garcia et al., 1995; Toonen & Verhage, 2007).

It works by binding *Syntaxin1* in closed conformation, allowing the assembly and stabilisation of the SNARE complex (Dulubova *et al.*, 2007; Toonen & Verhage, 2007). Missense mutations and deletion-derived haploinsufficiency of *Stxbp1* were first associated with epileptic encephalopathies in 2008 (Saitsu *et al.*, 2008). *Stxbp1* has since then become one of the genes most associated with epileptic encephalopathies in new-borns and children, and linked to movement and behavioural disorders (Khaikin & Mercimek-Mahmutoglu, 2016).

Null mice derived from deletion of both *Stxbp1* alleles result in viable embryos and develop normally assembled brains with fully differentiated neurons that produce spontaneous action potentials. These neurons though are then unable to respond to synaptic stimuli because the lack of *Stxbp1* makes them incapable of neurotransmitter release. The neurons undergo apoptosis after E12, and pups die at birth (Verhage *et al.*, 2000). The same developmentally healthy but eventually apoptotic phenotype was observed in human neurons *in vitro*: loss of function *Stxbp1* mutations in hESC-derived neurons were sufficient to induce neuronal death, but did not impact ESC to neuron specification (Patzke *et al.*, 2015).

Based on *Stxbp1*'s role in neuronal survival, and my observations upon *Stxbp1* KO, I wanted to verify whether *Stxbp1* was having the same role in iNs.

RNAseq (Treutlein *et al.*, 2016) shows it is selectively expressed at progressively higher levels in converting and maturing iNs (Fig. 5.5a). The reduction in iN numbers upon KO was confirmed by TUBB3 staining, as shown by whole well images and quantifications of iNs in Fig.5.5b-c, and shows again as a 50% conversion efficiency decrease (for both gRNAs chosen for analysis).

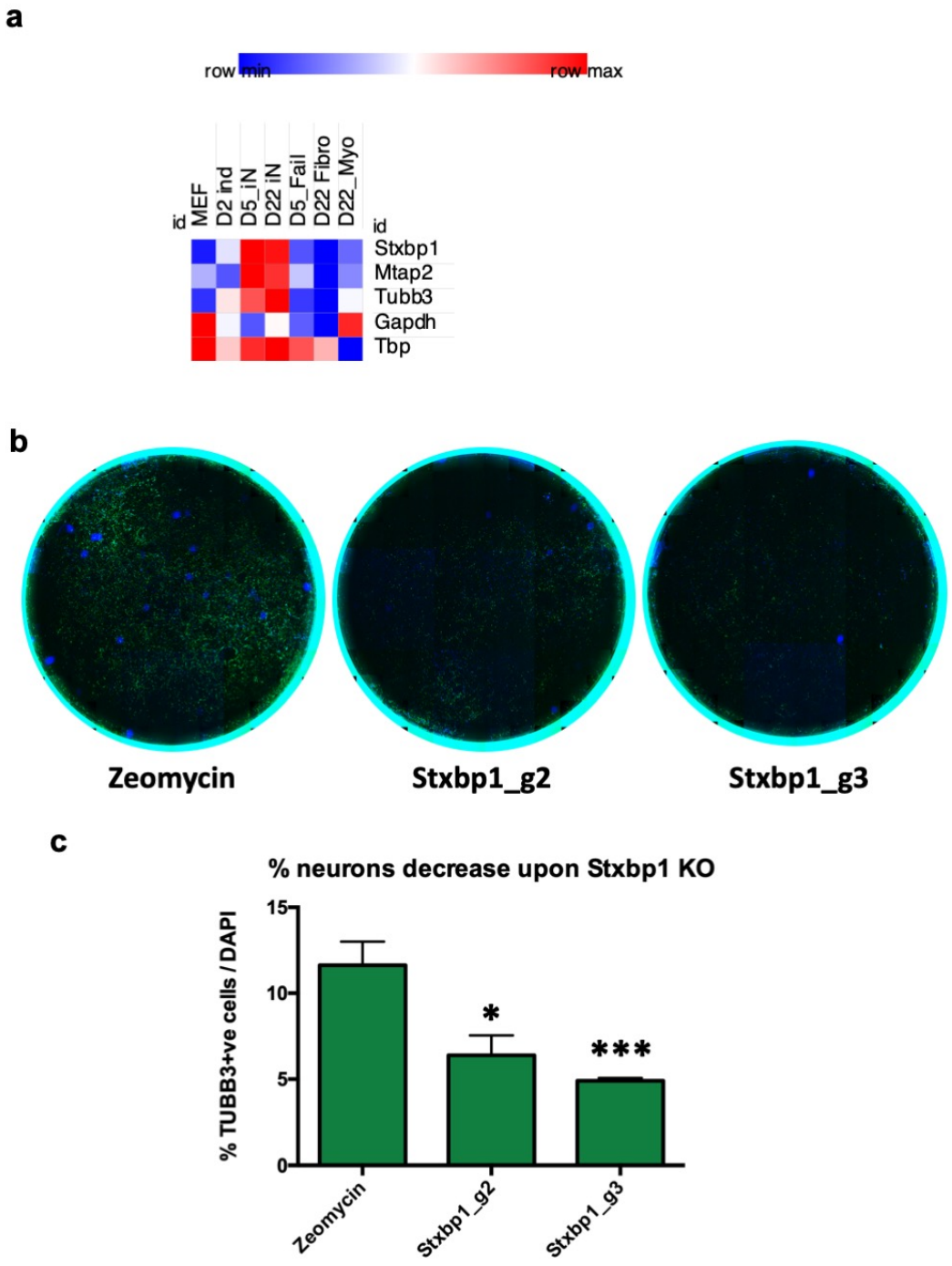


Fig. 5.5 – Stxbp1 expression across the conversion and KO effect by IF. (a) Gene expression of *Stxbp1* through conversion – data from (Treutlein *et al.*, 2016). Neuronal markers *Tubb3* and *Map2*, and housekeeping genes *Gapdh* and *Tbp* are shown as reference. **(b)** Whole well images of day 15 transdifferentiating cells upon transduction of neutral (*Pecam1*) gRNA or *Stxbp1* gRNAs #2 and #3. IF performed against TUBB3 shows visually a severe decrease in iNs that does not impact the overall cell numbers – based on DAPI stain.

(c) Quantification of b. *Stxbp1* KO results in about 50% depletion in TUBB3+ve cells, consistent with 50% depletion of dsRED+ve cells as reported by flow cytometry (n=2). Error bars indicate standard deviation. ***p<0.005, **p<0.01, *p<0.05 based on an unpaired two-tailed Student's t test.

To verify whether this reduction was coming from decreased neuronal specification or post-specification neuronal death, I monitored the % of iNs – as assessed by dsRED+ve cells with flow cytometry – from day 6, the earliest time-point of dsRED signal detection.

As shown in Fig.5.6a, at day 6 the percentage of dsRED +ve cells is the same in the control sample (using *Pecam1* gRNA) and the samples that received *Stxbp1* gRNAs #2 or #3, indicating that *Stxbp1* KO is not having an impact on induction of neuronal fate, consistent with reported phenotypes of *in vivo* neuronal differentiation of *Stxbp1 null* mice.

By day 10, *Stxbp1* KO samples show a significant decrease in dsRED +ve cells compared to control, already close to the 50% decrease reached at day 15. This suggests that neuronal cells start to die off soon after conversion, indicating the fundamental role of *Stxbp1* in early iN survival and consistent with the reported findings that *Stxbp1* KO neurons die off when synaptic activity would normally emerge.

To confirm that cell death is specific to neuronal cells only, cell counts for whole wells at day 6, 10 and 15 (Fig.5.6b) show that the overall number of cells is not being affected by the KO. This is consistent with the selective expression of *Stxbp1* in neurons, as cells lacking its expression are not likely to require it for survival, and consistent with null mice showing brain-specific defects, as well as mutations in *Stxbp1* in humans resulting in neurological disorders specifically.

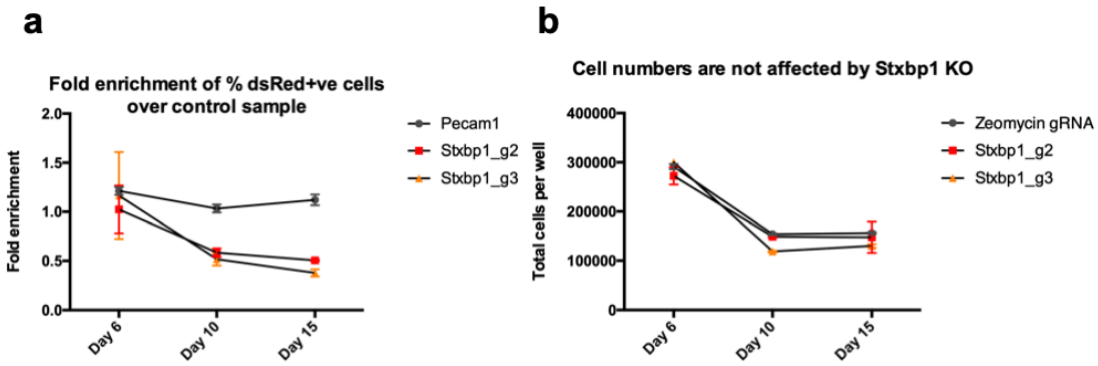


Fig. 5.6 – Effects of Stxbp1 KO on iNs and overall cell numbers across the early stages of conversion. (a) Fold enrichment of iNs (assessed as *hSyn1*-dsRed+ve cells by flow cytometry) over the control (Zeomycin) gRNA of cells transduced with the intronic gRNA against Zeomycin, or *Stxbp1* gRNAs #2 and #3. The number of iNs is similar – and more variable – at day 6, but then stabilises and shows as early as day 10 a marked decrease upon *Stxbp1* KO. (n=3). **(b)** The same gRNAs show no effect on overall cell growth, as the total number of cells at day 6, 10 and 15 is unvaried upon *Stxbp1* KO compared to the non-targeting gRNA (Zeomycin) (n=2). Error bars indicate standard deviation.

5.2.6 *Stxbp1* overexpression does not induce an increase in converted iNs

Once I verified that the phenotype of *Stxbp1* KO was consistent with the gene's reported role in neurons, I looked at how overexpressing it might affect neuronal physiology.

Overexpression of *Stxbp1* in *Drosophila* leads to inhibition of neurotransmitter release, possibly due to *Stxbp1* sequestering *Syntaxin1* and impairing SNARE complex formation and vesicle release (Schulze et al., 1994; Wu et al., 1998).

Stxbp1 has also been found to be upregulated in patients affected by schizophrenia (Behan et al., 2009).

Stxbp1 has two isoforms. Isoform A is longer and expressed selectively in the brain and retina. Isoform B is predominantly expressed in the cerebellum and retina, but can be found at low levels ubiquitously in rats and humans (Swanson, Steel & Valle, 1998).

In order to generate a mouse model for schizophrenia, isoform A was selectively overexpressed in glutamatergic and GABAergic neurons. This indeed led to schizophrenic-like cellular, molecular and behavioural changes, consistent with data from human patients indicating that neuroexocytosis defects related to *Stxbp*1 could be key regulators of schizophrenia (Urigüen *et al.*, 2013).

In the iN system, *Stxbp1* is upregulated early after *Ascl1* induction (levels are increased at day 2 already), at the same time as cells are exiting the cell cycle and switching to neuronal programming. I decided to test the effects of overexpressing *Stxbp1* on the conversion. I hypothesised that the increased presence of *Stxbp1* might influence the membrane properties of cells, or influence their vesicle release, and possibly push partially-reprogrammed cells to fully upregulate neuronal markers. I also hypothesised that overexpression of *Stxbp1* might influence the vesicle trafficking properties and hence firing properties of early neurons, which in turn could influence the behaviour/identity of their neighbouring cells.

To test these hypotheses, I cloned the cDNAs of both isoforms of *Stxbp1* into the tetO lentiviral backbone used for the expression of the Wernig factors.

I decided to test both isoforms. The cDNAs were PCR-amplified from whole cDNA extracted from adult mouse brain. Primers contain EcoRI restriction sites to allow ligation into the FUW-TetO backbone.

The PCR products were processed with gel electrophoresis and correctly showed a single band at the expected length of 1.6kb.

After purification, PCR products were digested and ligated within the pre-digested (EcoRI) tetO lentiviral backbone. The orientation of the insert was verified with restriction enzyme digestion, and the sequence of each isoform was confirmed via Sanger sequencing.

The overexpression construct was transduced into transdifferentiating MEFs together with the Wernig factors via lentiviruses. As an overexpression control, a

vector expressing BFP (Blue fluorescent protein) under the control of the tetO promoter (cloned within the same FUW-TetO backbone) was used.

Converting MEFs were transduced at day -1 with tetO-*BFP* or tetO-*Stxbp1A*/*B* at MOI5 together with the BAM factors. Dox was added at day 0 to induce TF expression. At day 3, cells were transduced with the *hSyn1*-dsRed construct to allow quantification of iNs.

Fig.5.7 shows the results of overexpressing both *Stxbp1* isoforms on day 15 iNs. Panel a shows the result of *Stxbp1* overexpression within the full BAM reprogramming cocktail. Panel b shows the result of overexpressing *Stxbp1* in *Ascl1*-only reprogramming – the idea being that *Stxbp1* OE might push more cells towards neuronal programming in the less efficient system.

As assessed by dsRED+ cells percentage, neither isoform impacts the neuronal yield of the conversion as compared to the BFP control sample.

This suggests that overexpression of *Stxbp1* in MEFs is not sufficient to push cells changing fate towards a neuronal lineage.

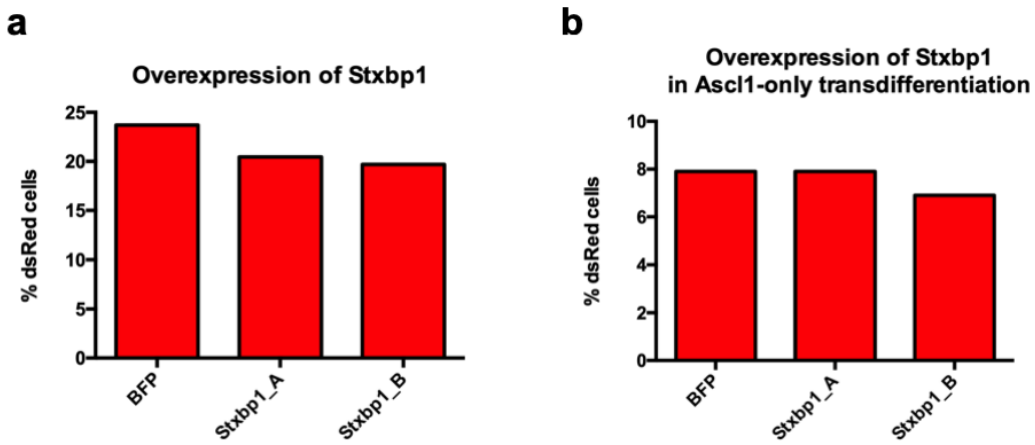


Fig. 5.7 – Overexpression of Stxbp1 does not influence iN conversion efficiency. (a) % of dsRED+ve cells at day 15 of the conversion upon overexpression of BFP, *Stxbp1* isoform A or *Stxbp1* isoform B. **(b)** % of *hSyn1*-dsRed+ve cells at day 15 of the conversion in the presence of *Ascl1* alone, upon overexpression of BFP, *Stxbp1* isoform A or *Stxbp1* isoform B. n=1.

It is reasonable to hypothesise that *Stxbp1* overexpression might be impacting membrane properties, neurotransmitter release and firing abilities of the iNs. Visual inspection did not show any clear difference in neuronal morphology or clustering between control and *Stxbp1* OE, not even in *Ascl1*-only reprogramming, but a more thorough investigation of the neurons' properties might reveal important differences. It would be for instance very interesting to test whether and how *Stxbp1* is impacting the different neuronal subtypes within the culture.

5.3 Discussion

5.3.1 *Zfp238* KO unexpectedly did not decrease iN conversion efficiency

As shown in Fig. 5.3, the knock-out of *Zfp238* with 3 different gRNAs in BAM-transduced *Cas9*-MEFs did not lead to a reduction in the number of dsRED+ve iNs compared to the control gRNA as assessed by flow cytometry. Of course, there is a possibility that all 3 gRNAs are inefficient at inducing indels, thus resulting in a functional ZFP238 protein that will not impact the conversion efficiency. The knock-out phenotype reported by Wapinski and colleagues though is quite striking: when reprogramming MEFs into iNs with *Ascl1* alone, they show an 8-fold decrease in the number of iNs generated from *Zfp238* KO MEFs as assessed by TUBB3 and MAP2 staining (Wapinski *et al.*, 2017). It seems unlikely that, shall that phenotype also verify in BAM-transduced MEFs, neither one of the 3 gRNAs were able to induce even a mild decrease in iN conversion efficiency.

An alternative explanation is that, considering the huge differences in efficiency between *Ascl1*-only and BAM-mediated iN reprogramming (Vierbuchen *et al.*, 2010; Wapinski *et al.*, 2013), the knock-out of *Zfp238* would severely decrease the efficiency of the former but not the latter. *Brn2* and *Myt1l* are known to work by stabilizing and maintaining neuronal fate induced mostly by *Ascl1* (as well as, especially *Myt1l*, improving the maturation of the iNs), and it's possible that their presence would be able to compensate for the absence of *Zfp238* in directing the

neuronal network. As mentioned, *Zfp238* and *Myt1l* can generate iNs in the absence of *Ascl1* (Wapinski *et al.*, 2013, 2017), indicating that the induction of neuronal programming can be achieved through compensatory mechanisms that upregulate the same shared network.

To verify this hypothesis, I could replicate the same experiment by transducing *Cas9*-MEFs with the *Zfp238* gRNAs and *Ascl1*-only lentiviruses, then assess the neuronal yield using *hSyn1-dsRed*. It would be useful to also assess neuronal conversion efficiency in parallel, for both *Ascl1*-only and BAM reprogramming, using TUBB3/MAP2 staining, to verify whether this phenotype discrepancy is coming from the different neuronal reporters used for quantification. As mentioned in the discussion of chapter 3 regarding *Tau* and *Tubb3* differences, it's possible that *hSyn1-dsRed* is detecting pro-neuronal cells in *Zfp238* KO conditions that do not upregulate full neuronal markers like *Tubb3* and *Map2* and would thus not be picked up by staining.

5.3.2 *Stxbp1* knock-out iNs as an *in vitro* model for neurodevelopmental disorders.

The results shown in Fig.5.5-6 confirm what is known about *Stxpb1*'s role in *in vivo* neurons: its expression is specific for neuronal cells, and its presence is required for neuronal survival.

I have not addressed in this thesis whether the iNs are going through apoptosis following the same pathway as their *in vivo* counterparts, but it is a reasonable assumption based on what is known about their gene expression and physiological similarities.

Models of *Stxbp1* haploinsufficiency in mice have shown enhanced aggressiveness and impaired fear-mediated learning, both of which can be ameliorated by artificially potentiating excitatory synaptic transmission (Miyamoto *et al.*, 2017). But haploinsufficiency is not the only possible cause for the neurological diseases associated with *Stxbp1* mutations: missense mutations can give rise to cellular aggregates of *Stxbp1* that impair its ability to form neuroexocytosis complexes

(Martin *et al.*, 2014). Recently, mutant aggregative forms of *Stxbp1* have been shown to aggregate wt forms of the protein as well in a prion-like manner, and induce the formation of *α-synuclein* aggregates and Lewy body-like structures in cultured rat hippocampal neurons (Chai *et al.*, 2016). This study showed how *Stxbp1* acts as a molecular chaperone for *α-synuclein*, how its haploinsufficiency leads to an increased *α-synuclein* aggregation phenotype, and potentially linked for the first time two apparently distinct neurodevelopmental and neurodegenerative disorders.

The versatility of the iN system makes it an ideal platform to study patient-derived, subtype specific neuronal defects arising from specific mutations. The work in this thesis on *Stxbp1* KO confirms it can faithfully reproduce *in vivo* phenotypes and lead to iN death. The integration of *Cas9*-expressing cells within the conversion system would potentially allow the introduction of multiple mutations in subtype specific cells with relative ease and efficiency, and possibly expand on the newly-discovered interactions between aggregating disease-associated proteins of neurological diseases.

5.3.3 *Stxbp1* overexpression – a candidate *in vitro* model for schizophrenia?

Regulation of *Stxbp1* expression levels seems to be critical for correct neuronal synaptic function. On the opposite end of haploinsufficiency, *Stxbp1* upregulation has been identified in human patients suffering from schizophrenia (Behan *et al.*, 2009) and *Stxbp1a* overexpression in mouse glutamatergic and GABAergic neurons induced in the animals schizophrenia-like behaviours (Urigüen *et al.*, 2013). Interestingly, this model also showed a decrease of D1 receptors of the dopamine signalling pathway and downregulation of the DA membrane transporter, as well as tyrosine hydroxylase overexpression in the cortex and striatum.

In this chapter I've tested how the overexpression of *Stxbp1* could affect the conversion efficiency of MEFs into iNs, using both the BAM combination and *Ascl1* alone. As assessed by *hSyn1-dsRed* quantification, the conversion efficiency is neither decreased nor increased upon *Stxbp1* overexpression. This is consistent with the

knock-out phenotype at day 5, suggesting that in the iN system *Stxbp1* levels do not impact the neuronal fate induction machinery. Expression levels of *Stxbp1* in iNs are comparable between day 5 and day 22, suggesting that the required amounts for proper neuronal function are expressed early on in neuronal specification. Considering what is known *about* Stxbp1 overexpression though, it is possible that the iNs overexpressing it would have differences in their neuroexocytosis properties. In such an environment, it would be interesting to test whether it might skew the relative abundance of neuronal subtypes, or impact the behaviour of some types more than others – for instance, it would be worth investigating whether overexpressing *Stxbp1* in induced dopaminergic neurons makes them more active, or increases their dopamine release, or impacts their receptor expression.

5.4 Summary

This chapter aimed to examine the role of 20 genes in the MEF to iN system using the CRISPR/Cas9-mediated KO to assess gene function, and *hSyn1*-dsRed expression to quantify neurons, both tools established in Chapter 3.

The first important result is that *Kmt2b* KO in this system leads to a decrease in the conversion efficiency as assessed by dsRED+ve iNs via flow cytometry. This confirms previously published results from *Km2tb null* and heterozygous mouse cells, showing that *Kmt2b* KO leads to a reduction in iNs as assessed by TUBB3 and PSA-NCAM expression. This further validation confirms that the platform discussed in this thesis is reliable and allows the investigation of genetic phenotypes in a complex system such as iNs quickly and easily, bypassing the need for KO mouse lines and allowing quantification of iNs with a simple and specific reporter.

The second important result is the discovery that *Stxbp1* KO decreases the yield of iN by 50%. This is not achieved by impairing the cell identity switch, but likely through neuronal apoptosis following neurotransmitter accumulation due to impaired vesicle

release. This is further proof of how iNs reproduce characteristics and behaviours of *in vivo* neurons despite their origin and heterogeneous cell environment, and strengthens the fibroblast to neuron conversion platform as a system to model neurological diseases, perform drug testing and potentially personalised medicine.

Chapter 6 Conclusion

In the 9 years since the establishment of the first induced neurons from mouse fibroblasts, a lot has been learned about their biology, their similarities and differences with *in vivo* neurons, their abilities to mature and form networks of synchronized firing. Different combinations of transcription factors, differentiation media, small molecule inducers and cells of origin can give rise to a huge variety of neuronal subtypes (reviewed in Colasante *et al.*, 2019). Though lots of efforts have gone into increasing the efficiency of the conversion and the neuronal purity of the end culture, the genetic and molecular mechanisms driving the reprogramming processes are still not well understood. RNAseq, ATACseq and ChIPseq analyses have contributed to broaden our knowledge of the gene regulatory network that suppresses fibroblast identity and promotes the specification of neuronal fate (Wapinski *et al.*, 2013, 2017; Treutlein *et al.*, 2016; Mall *et al.*, 2017), but how that network differs between cell sources and neuronal subtypes, or If and how it can be manipulated to optimise the different systems still remains to be addressed.

The iN system also represents an interesting model to study cell identity conversions between somatic cell states, particularly because it does not pass through a stem-cell intermediate (Vierbuchen *et al.*, 2010; Treutlein *et al.*, 2016), and thus does not mimic differentiation mechanisms of cell specification. It would thus represent a suitable platform to investigate processes such as genome re-organisation and chromatin remodelling that have been broadly associated and implicated in fate specifications and conversions.

One major remaining caveat in this system is the heterogenicity of the cultures. Although purity, even within specific neuronal subtypes, has been significantly enhanced in the last few years, the intrinsic origin of the cells, i.e. the non-neural environment that iNs are subjected to – be it fibroblast and myoblast-like cells, or blood lineages, exposes the cells to very non-physiological conditions that are bound to affect their properties. In fact, most publications that focus on neuronal

maturation and functionality, for which electrophysiology is the gold standard, rely on iN sorting followed by co-culture with *ex-vivo* astrocytes or cortical neurons (Vierbuchen *et al.*, 2010; Caiazzo *et al.*, 2011; Pfisterer *et al.*, 2011; Pereira *et al.*, 2014; Ruetz *et al.*, 2017; reviewed in Colasante *et al.*, 2019) . In this regard, astrocyte-derived iNs, both *in vitro* and especially *in vivo*, represent perhaps a better source for maturation of iNs, as they also provide a more nurturing environment. The advantage of *in vivo* iN reprogramming is that the neurons can directly integrate with the local circuitry and respond to endogenous physiological stimuli.

The system established in this thesis represents a strong platform to perform loss-of-function studies of specific target genes in iNs in a quick and easy way. Constitutive *Cas9* expression allows the temporal control of gene knock-out, achieved efficiently by gRNA lentiviral transduction. This means that KO could be directly induced within neurons, bypassing the need to engineer murine animal models to study the genetic and molecular pathology of neurological diseases.

Subtype-specific promoters could be used to drive gRNA expression and allow KO within selective populations. This could be used as a strategy to understand, even within a heterogeneous context, how selected genes act in selected subtypes.

The availability of mouse lines that constitutively and ubiquitously express *Cas9* would also allow the investigation of knock-out phenotypes in *in vivo* converted iNs by lentiviral delivery of gRNAs in target brain regions.

The *hSyn1* promoter activity documented in this thesis could bypass the need to engineer mouse models by crossing them with the *Tau*-eGFP line to make use of that reporter to mark iNs. At the same time, it bypasses the need of using antibodies to sort neurons based on the expression of specific extracellular antigens like PSA-NCAM. Considering neuronal yield, the loss of cells associated with staining procedures for flow cytometry, and the promiscuity of certain antibodies, this represents a significant advantage.

Although the screening performed was unsuccessful, the tools and knowledge developed would be useful in optimising the conditions and potentially perform a more targeted, stringent and higher-coverage screen. *Kmt2b*, *Sf3a1* and *Stxbp1* KO phenotypes verified in this thesis would serve as controls in setting-up an efficient platform with higher chances of success.

The work performed on *Stxbp1* confirmed the validity of iNs as modelling platform for their *in vivo* counterpart and their suitability to mimic disease phenotypes.

The discovery of *Sf3a1* as an essential regulator of cell fitness in this context, as well as modulator of the iN conversion efficiency, highlights the role of the relatively unexplored area of RNA processing in cell identity conversions.

Overall, this work has established a solid platform for characterising induced neurons and has given examples of how it can reveal basic biology pathways as well as reproduce disease phenotypes. A new and improved screen plan could reveal insights into the genetic machinery that controls iN specification and improve its suitability for future personalised medicine applications.

Chapter 7 References

Adler, A.F., Grigsby, C.L., Kulangara, K., Wang, H., et al. (2012) Nonviral Direct Conversion of Primary Mouse Embryonic Fibroblasts to Neuronal Cells. *Molecular Therapy - Nucleic Acids*. [Online] 1, e32. Available from: doi:10.1038/MTNA.2012.25.

Almendros, C., Kieper, S.N. & Brouns, S.J.J. (2019) CRISPR-Cas Systems Reduced to a Minimum. *Molecular Cell*. [Online] 73 (4), 641–642. Available from: doi:10.1016/J.MOLCEL.2019.02.005.

Anon (2019) The Gene Ontology Resource: 20 years and still GOing strong. [Online] 47 (D1), D330–D338. Available from: doi:10.1093/nar/gky1055.

De Arras, L. & Alper, S. (2013) Limiting of the innate immune response by SF3A-dependent control of MyD88 alternative mRNA splicing. *PLoS genetics*. [Online] 9 (10), e1003855. Available from: doi:10.1371/journal.pgen.1003855.

Barbagiovanni, G., Germain, P.-L., Zech, M., Atashpaz, S., et al. (2018) KMT2B Is Selectively Required for Neuronal Conversion, and Its Loss Exposes Dystonia Candidate Genes. *Cell Reports*. [Online] 25 (4), 988–1001. Available from: doi:10.1016/J.CELREP.2018.09.067.

Barker, R.A., Götz, M. & Parmar, M. (2018) New approaches for brain repair—from rescue to reprogramming. *Nature*. [Online] 557 (7705), 329–334. Available from: doi:10.1038/s41586-018-0087-1.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., et al. (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*. [Online] 315 (5819), 1709–1712. Available from: doi:10.1126/science.1138140.

Behan, Á., Byrne, C., Dunn, M.J., Cagney, G., et al. (2009) Proteomic analysis of membrane microdomain-associated proteins in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder reveals alterations in LAMP, STXBP1 and BASP1 protein expression. *Molecular Psychiatry*. [Online] 14 (6), 601–613. Available from: doi:10.1038/mp.2008.7.

Bertrand, N., Castro, D.S. & Guillemot, F. (2002) Proneural genes and the specification of neural cell types. *Nature Reviews Neuroscience*. [Online] 3 (7), 517–530. Available from: doi:10.1038/nrn874.

Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*. 161 (3), 1169–1175.

Black, J.B., Adler, A.F., Wang, H.-G., DIppolito, A.M., et al. (2016) Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Stem Cell*. [Online] 19, 406–414. Available from: doi:10.1016/j.stem.2016.07.001.

Blau, H.M., Chiu, C.-P. & Webster, C. (1983) Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell*. [Online] 32 (4), 1171–1180. Available from: doi:10.1016/0092-8674(83)90300-8.

Bogen, I.L., Jensen, V., Hvalby, O. & Walaas, S.I. (2009) Synapsin-dependent development of glutamatergic synaptic vesicles and presynaptic plasticity in postnatal mouse brain. *Neuroscience*. [Online] 158 (1), 231–241. Available from: doi:10.1016/J.NEUROSCIENCE.2008.05.055.

Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal

origin. *Microbiology*. [Online] 151 (8), 2551–2561. Available from: doi:10.1099/mic.0.28048-0.

Boutros, M. & Ahringer, J. (2008) The art and design of genetic screens: RNA interference. *Nature Reviews Genetics*. [Online] 9 (7), 554–566. Available from: doi:10.1038/nrg2364.

Briggs, R. & King, T.J. (1952) Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 38 (5), 455–463. Available from: doi:10.1073/pnas.38.5.455.

Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*. [Online] 42 (22), e168–e168. Available from: doi:10.1093/nar/gku936.

Caiazzo, M., Dell'Anno, M.T., Dvoretskova, E., Lazarevic, D., et al. (2011) Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature*. [Online] 476 (7359), 224–227. Available from: doi:10.1038/nature10284.

Chai, Y.J., Sierecki, E., Tomatis, V.M., Gormal, R.S., et al. (2016) Munc18-1 is a molecular chaperone for α-synuclein, controlling its self-replicating aggregation. *The Journal of cell biology*. [Online] 214 (6), 705–718. Available from: doi:10.1083/jcb.201512016.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., et al. (2009) Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology*. [Online] 27 (3), 275–280. Available from: doi:10.1038/nbt.1529.

Chanda, S., Ang, C.E., Davila, J., Pak, C., et al. (2014) Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. *Stem Cell Reports*. [Online] 3 (2), 282–296. Available from: doi:10.1016/j.stemcr.2014.05.020.

Chanda, S., Marro, S., Wernig, M. & Sudhof, T.C. (2013) Neurons generated by direct conversion of fibroblasts reproduce synaptic phenotype caused by autism-associated neuroligin-3 mutation. *Proceedings of the National Academy of Sciences*. [Online] 110 (41), 16622–16627. Available from: doi:10.1073/pnas.1316240110.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., et al. (2017) Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*. [Online] 550 (7676), 407–410. Available from: doi:10.1038/nature24268.

Chen, X., Du, H., Liu, B., Zou, L., et al. (2015) The Associations between RNA Splicing Complex Gene SF3A1 Polymorphisms and Colorectal Cancer Risk in a Chinese Population. *PloS one*. [Online] 10 (6), e0130377. Available from: doi:10.1371/journal.pone.0130377.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*. [Online] 186 (2), 757–761. Available from: doi:10.1534/genetics.110.120717.

Colasante, G., Rubio, A., Massimino, L. & Broccoli, V. (2019) Direct Neuronal Reprogramming Reveals Unknown Functions for Known Transcription Factors. *Frontiers in Neuroscience*. [Online] 13, 283. Available from: doi:10.3389/fnins.2019.00283.

Cong, L., Ran, F.A., Cox, D., Lin, S., et al. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. [Online] 339 (6121), 819–823. Available from: doi:10.1126/science.1231143.

Davis, R.L., Weintraub, H. & Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*. [Online] 51 (6), 987–1000. Available from: doi:10.1016/0092-8674(87)90585-x.

Demaret, J., Gossez, M., Venet, F. & Monneret, G. (2017) *Intracellular Flow Cytometry Improvements in Clinical Studies*. In: [Online]. Humana Press, New York, NY. pp. 315–327. Available from: doi:10.1007/978-1-4939-6603-5_20.

Díaz-Hernández, J.I., Sebastián-Serrano, Á., Gómez-Villafuertes, R., Díaz-Hernández, M., et al. (2015) Age-Related Nuclear Translocation of P2X6 Subunit Modifies Splicing Activity Interacting with Splicing Factor 3A1 Barbara Bardoni (ed.). *PLOS ONE*. [Online] 10 (4), e0123121. Available from: doi:10.1371/journal.pone.0123121.

Doench, J.G. (2018) Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics*. [Online] 19 (2), 67–80. Available from: doi:10.1038/nrg.2017.97.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*. [Online] 34 (2), 184–191. Available from: doi:10.1038/nbt.3437.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., et al. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development*. 123 (1).

Drouin-Ouellet, J., Lau, S., Brattås, P.L., Rylander Ottosson, D., et al. (2017) REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways. *EMBO molecular medicine*. [Online] 9 (8), 1117–1131. Available from: doi:10.15252/emmm.201607471.

Dulubova, I., Khvotchev, M., Liu, S., Huryeva, I., et al. (2007) Munc18-1 binds directly

to the neuronal SNARE complex. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 104 (8), 2697–2702. Available from: doi:10.1073/pnas.0611318104.

Ed Yong (2016) A Google Maps For The Human Body - The Atlantic. *The Atlantic*.

Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., et al. (2019) A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Molecular Cell*. [Online] 73 (4), 714-726.e4. Available from: doi:10.1016/J.MOLCEL.2018.12.003.

Elling, U., Taubenschmid, J., Wirnsberger, G., O'Malley, R., et al. (2011) Forward and Reverse Genetics through Derivation of Haploid Mouse Embryonic Stem Cells. *Cell Stem Cell*. [Online] 9 (6), 563–574. Available from: doi:10.1016/j.stem.2011.10.012.

Evers, B., Jastrzebski, K., Heijmans, J.P.M., Grernrum, W., et al. (2016) CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature Biotechnology*. [Online] 34 (6), 631–633. Available from: doi:10.1038/nbt.3536.

Ferguson, E.L. & Horvitz, H.R. (1985) Identification and characterization of 22 genes that affect the vulval cell lineages of the nematode Caenorhabditis elegans. *Genetics*. 110 (1), 17–72.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., et al. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. [Online] 391 (6669), 806–811. Available from: doi:10.1038/35888.

Fishman, V., Shnayder, T., Orishchenko, K., Bader, M., et al. (2015) Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells. *Cell Cycle*. [Online] 14 (8), 1188–1196. Available from: doi:10.1080/15384101.2015.1012875.

Friedel, R.H. & Soriano, P. (2010) Gene Trap Mutagenesis in the Mouse. *Methods in Enzymology*. [Online] 477, 243–269. Available from: doi:10.1016/S0076-6879(10)77013-0.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., et al. (2011) An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell*. [Online] 147 (1), 132–146. Available from: doi:10.1016/J.CELL.2011.08.023.

Garcia, E.P., Gatti, E., Butler, M., Burton, J., et al. (2003) A rat brain Secl homologue related to Rop and UNC18 interacts with syntaxin (synaptic vesicle/exocytosis/VAMP/synaptobrevin/SNAP-25). *Proc. Nad. Acad. Sci. USA*.91.

Garcia, E.P., McPherson, P.S., Chilcote, T.J., Takei, K., et al. (1995) rbSec1A and B colocalize with syntaxin 1 and SNAP-25 throughout the axon, but are not in a stable complex with syntaxin. *The Journal of cell biology*. [Online] 129 (1), 105–120. Available from: doi:10.1083/jcb.129.1.105.

Gascón, S., Murenu, E., Masserdotti, G., Ortega, F., et al. (2016) Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. *Cell Stem Cell*. [Online] 18 (3), 396–409. Available from: doi:10.1016/J.STEM.2015.12.003.

Guo, Z., Zhang, L., Wu, Z., Chen, Y., et al. (2014) In Vivo Direct Reprogramming of Reactive Glial Cells into Functional Neurons after Brain Injury and in an Alzheimer's Disease Model. *Cell Stem Cell*. [Online] 14 (2), 188–202. Available from: doi:10.1016/J.STEM.2013.12.001.

Gurdon, J.B. (1962) Adult frogs derived from the nuclei of single somatic cells. *Developmental Biology*. [Online] 4 (2), 256–273. Available from: doi:10.1016/0012-

1606(62)90043-X.

Haffter, P., Granato, M., Brand, M., Mullins, M.C., et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development (Cambridge, England)*. 123, 1–36.

Han, H., Irimia, M., Ross, P.J., Sung, H.-K., et al. (2013) MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature*. [Online] 498 (7453), 241–245. Available from: doi:10.1038/nature12270.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., et al. (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. [Online] 163 (6), 1515–1526. Available from: doi:10.1016/j.cell.2015.11.015.

Hartwell, L.H., Culotti, J. & Reidt, B. (1970) Genetic Control of the Cell-Division Cycle in Yeast, I. Detection of Mutants. *Proceedings of the National Academy of Sciences*.66 (2).

Hata, A. & Chen, Y.-G. (2016) TGF-β Signaling from Receptors to Smads. *Cold Spring Harbor perspectives in biology*. [Online] 8 (9), a022061. Available from: doi:10.1101/cshperspect.a022061.

Hata, Y., Slaughter, C.A. & Südhof, T.C. (1993) Synaptic vesicle fusion complex contains unc-18 homologue bound to syntaxin. *Nature*. [Online] 366 (6453), 347–351. Available from: doi:10.1038/366347a0.

Hicks, G.G., Shi, E., Li, X.-M., Li, C.-H., et al. (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nature Genetics*. [Online] 16 (4), 338–344. Available from: doi:10.1038/ng0897-338.

Hochedlinger, K. & Jaenisch, R. (2002) Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature*. [Online] 415 (6875), 1035–1038. Available from: doi:10.1038/nature718.

Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome Biology*. [Online] 14 (10), R115. Available from: doi:10.1186/gb-2013-14-10-r115.

Hsu, P.D., Lander, E.S. & Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. [Online] 157 (6), 1262–1278. Available from: doi:10.1016/j.cell.2014.05.010.

Hu, W., Qiu, B., Guan, W., Wang, Q., et al. (2015) Direct Conversion of Normal and Alzheimer's Disease Human Fibroblasts into Neuronal Cells by Small Molecules. *Cell Stem Cell*. [Online] 17 (2), 204–212. Available from: doi:10.1016/J.STEM.2015.07.006.

Huh, C.J., Zhang, B., Victor, M.B., Dahiya, S., et al. (2016) Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. *eLife*. [Online] 5. Available from: doi:10.7554/eLife.18648.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., et al. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*. [Online] 169 (12), 5429–5433. Available from: doi:10.1128/jb.169.12.5429-5433.1987.

Jackson, A.L. & Linsley, P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery*. [Online] 9 (1), 57–67. Available from: doi:10.1038/nrd3010.

Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., et al. (2017) CRISPR-Cas:

Adapting to change. *Science (New York, N.Y.)*. [Online] 356 (6333), eaal5056. Available from: doi:10.1126/science.aal5056.

Jansen, R., Embden, J.D.A. van, Gaastra, W. & Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*. [Online] 43 (6), 1565–1575. Available from: doi:10.1046/j.1365-2958.2002.02839.x.

Jiang, H., Xu, Z., Zhong, P., Ren, Y., et al. (2015) Cell cycle and p53 gate the direct conversion of human fibroblasts to dopaminergic neurons. *Nature Communications*. [Online] 6 (1), 10100. Available from: doi:10.1038/ncomms10100.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*. [Online] 337 (6096), 816–821. Available from: doi:10.1126/science.1225829.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., et al. (2017) Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*. [Online] 12 (4), 828–863. Available from: doi:10.1038/nprot.2017.016.

Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., et al. (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nature neuroscience*. [Online] 18 (9), 1226–1229. Available from: doi:10.1038/nn.4085.

Kalidas, S., Santosh, V., Shareef, M.M., Shankar, S.K., et al. (2000) Expression of p67 (Munc-18) in adult human brain and neuroectodermal tumors of human central nervous system. *Acta Neuropathologica*. [Online] 99 (2), 191–198. Available from: doi:10.1007/PL00007424.

Khaikin, Y. & Mercimek-Mahmutoglu, S. (2016) *STXBP1 Encephalopathy with*

Epilepsy. University of Washington, Seattle.

Kim, J., Efe, J.A., Zhu, S., Talantova, M., et al. (2011) Direct reprogramming of mouse fibroblasts to neural progenitors. *Proceedings of the National Academy of Sciences*. [Online] 108 (19), 7838–7843. Available from: doi:10.1073/pnas.1103113108.

Knight, S.C., Xie, L., Deng, W., Guglielmi, B., et al. (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science (New York, N.Y.)*. [Online] 350 (6262), 823–826. Available from: doi:10.1126/science.aac6572.

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M.D.C., et al. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*. [Online] 32 (3), 267–273. Available from: doi:10.1038/nbt.2800.

Kondo, M., Scherer, D.C., Miyamoto, T., King, A.G., et al. (2000) Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines. *Nature*. [Online] 407 (6802), 383–386. Available from: doi:10.1038/35030112.

Krämer, A., Mulhauser, F., Wersig, C., Gröning, K., et al. (1995) Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of Saccharomyces cerevisiae. *RNA (New York, N.Y.)*. 1 (3), 260–272.

Kügler, S., Kilic, E. & Bähr, M. (2003) Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Therapy*. [Online] 10 (4), 337–347. Available from: doi:10.1038/sj.gt.3301905.

Kweon, J. & Kim, Y. (2018) High-throughput genetic screens using CRISPR–Cas9

system. *Archives of Pharmacal Research*. [Online] 41 (9), 875–884. Available from: doi:10.1007/s12272-018-1029-z.

Lancaster, M.A. & Knoblich, J.A. (2014) Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science*. [Online] 345 (6194), 1247125–1247125. Available from: doi:10.1126/science.1247125.

Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., et al. (2013) Cerebral organoids model human brain development and microcephaly. *Nature*. [Online] 501 (7467), 373–379. Available from: doi:10.1038/nature12517.

Leeb, M., Dietmann, S., Paramor, M., Niwa, H., et al. (2014) Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell stem cell*. [Online] 14 (3), 385–393. Available from: doi:10.1016/j.stem.2013.12.008.

Li, W., Xu, H., Xiao, T., Cong, L., et al. (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*. [Online] 15 (12), 554. Available from: doi:10.1186/s13059-014-0554-4.

Li, X., Zuo, X., Jing, J., Ma, Y., et al. (2015) Small-Molecule-Driven Direct Reprogramming of Mouse Fibroblasts into Functional Neurons. *Cell Stem Cell*. [Online] 17 (2), 195–203. Available from: doi:10.1016/J.STEM.2015.06.003.

Lietz, M., Cicchetti, P. & Thiel, G. (1998) Inverse expression pattern of REST and synapsin I in human neuroblastoma cells. *Biological chemistry*. 379 (10), 1301–1304.

Liu, M.-L., Zang, T. & Zhang, C.-L. (2016) Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients. *Cell reports*. [Online] 14 (1), 115–128. Available from: doi:10.1016/j.celrep.2015.12.018.

Liu, Y., Yu, C., Daley, T.P., Wang, F., et al. (2018) CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell*. [Online] 23 (5), 758-771.e8. Available from: doi:10.1016/J.STEM.2018.09.003.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., et al. (2008) Highly parallel identification of essential genes in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 105 (51), 20380–20385. Available from: doi:10.1073/pnas.0810485105.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., et al. (2011) Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*. [Online] 9 (6), 467–477. Available from: doi:10.1038/nrmicro2577.

Makarova, K.S., Zhang, F. & Koonin, E. V (2017) *SnapShot: Class 2 CRISPR-Cas Systems*. [Online] Available from: doi:10.1016/j.cell.2016.12.038.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., et al. (2013) RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*. [Online] 339 (6121), 823–826. Available from: doi:10.1126/science.1232033.

Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., et al. (2017) Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature*. [Online] 544 (7649), 245–249. Available from: doi:10.1038/nature21722.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.-C., et al. (2011) Direct Lineage Conversion of Terminally Differentiated Hepatocytes to Functional Neurons. *Cell Stem Cell*. [Online] 9 (4), 374–382. Available from: doi:10.1016/J.STEM.2011.09.002.

Martin, S., Papadopulos, A., Tomatis, V.M., Sierecki, E., et al. (2014) Increased

Polyubiquitination and Proteasomal Degradation of a Munc18-1 Disease-Linked Mutant Causes Temperature-Sensitive Defect in Exocytosis. *Cell Reports*. [Online] 9 (1), 206–218. Available from: doi:10.1016/J.CELREP.2014.08.059.

Maryanovich, M. & Gross, A. (2013) A ROS rheostat for cell fate regulation. *Trends in Cell Biology*. [Online] 23 (3), 129–134. Available from: doi:10.1016/J.TCB.2012.09.007.

Masserdotti, G., Gascón, S. & Götz, M. (2016) Direct neuronal reprogramming: learning from and for development. *Development*. [Online] 143 (14), 2494–2510. Available from: doi:10.1242/dev.092163.

McLean, J.R., Smith, G.A., Rocha, E.M., Hayes, M.A., et al. (2014) Widespread neuron-specific transgene expression in brain and spinal cord following synapsin promoter-driven AAV9 neonatal intracerebroventricular injection. *Neuroscience Letters*. [Online] 576, 73–78. Available from: doi:10.1016/J.NEULET.2014.05.044.

Mertens, J., Paquola, A.C.M., Ku, M., Hatch, E., et al. (2015) Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell*. [Online] 17 (6), 705–718. Available from: doi:10.1016/j.stem.2015.09.001.

Mertens, J., Reid, D., Lau, S., Kim, Y., et al. (2018) Aging in a Dish: iPSC-Derived and Directly Induced Neurons for Studying Brain Aging and Age-Related Neurodegenerative Diseases. *Annual review of genetics*. [Online] 52, 271–293. Available from: doi:10.1146/annurev-genet-120417-031534.

Meshorer, E. & Soreq, H. (2002) Pre-mRNA splicing modulations in senescence. *Aging Cell*. [Online] 1 (1), 10–16. Available from: doi:10.1046/j.1474-9728.2002.00005.x.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., et al. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*. [Online] 47 (D1), D419–D426. Available from: doi:10.1093/nar/gky1038.

Miller, J.D., Ganat, Y.M., Kishinevsky, S., Bowman, R.L., et al. (2013) Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell stem cell*. [Online] 13 (6), 691–705. Available from: doi:10.1016/j.stem.2013.11.006.

Miyamoto, H., Shimohata, A., Abe, M., Abe, T., et al. (2017) Potentiation of excitatory synaptic transmission ameliorates aggression in mice with Stxbp1 haploinsufficiency. *Human Molecular Genetics*. [Online] 26 (24), 4961–4974. Available from: doi:10.1093/hmg/ddx379.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., et al. (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science (New York, N.Y.)*. [Online] 353 (6299), aad5147. Available from: doi:10.1126/science.aad5147.

Mohr, S., Bakal, C. & Perrimon, N. (2010) Genomic screening with RNAi: results and challenges. *Annual review of biochemistry*. [Online] 79, 37–64. Available from: doi:10.1146/annurev-biochem-060408-092949.

Mohr, S.E., Smith, J.A., Shamu, C.E., Neumüller, R.A., et al. (2014) RNAi screening comes of age: improved techniques and complementary approaches. *Nature Reviews Molecular Cell Biology*. [Online] 15 (9), 591–600. Available from: doi:10.1038/nrm3860.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system.

Microbiology. [Online] 155 (3), 733–740. Available from: doi:10.1099/MIC.0.023960-0.

Mojica, F.J.M., Diez-Villasenor, C., Soria, E. & Juez, G. (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*. [Online] 36 (1), 244–246. Available from: doi:10.1046/j.1365-2958.2000.01838.x.

Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. (2005) Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*. [Online] 60 (2), 174–182. Available from: doi:10.1007/s00239-004-0046-3.

Nord, A.S., Chang, P.J., Conklin, B.R., Cox, A. V., et al. (2006) The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Research*. [Online] 34 (90001), D642–D648. Available from: doi:10.1093/nar/gkj097.

Nurse, P. (1975) Genetic control of cell size at cell division in yeast. *Nature*. [Online] 256 (5518), 547–551. Available from: doi:10.1038/256547a0.

Nusse, R. & Clevers, H. (2017) *Wnt/b-Catenin Signaling, Disease, and Emerging Therapeutic Modalities*. [Online] Available from: doi:10.1016/j.cell.2017.05.016.

Nüsslein-Volhard, C. & Wieschaus, E. (1980) Mutations affecting segment number and polarity in Drosophila. *Nature*. [Online] 287 (5785), 795–801. Available from: doi:10.1038/287795a0.

O'Connor, B.P., Danhorn, T., De Arras, L., Flatley, B.R., et al. (2015) Regulation of Toll-like Receptor Signaling by the SF3a mRNA Splicing Complex Xin Lin (ed.). *PLOS*

Genetics. [Online] 11 (2), e1004932. Available from: doi:10.1371/journal.pgen.1004932.

Ohta, S., Nishida, E., Yamanaka, S. & Yamamoto, T. (2013) Global Splicing Pattern Reversion during Somatic Cell Reprogramming. *Cell Reports*. [Online] 5 (2), 357–366. Available from: doi:10.1016/J.CELREP.2013.09.016.

Okita, K., Ichisaka, T. & Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells. *Nature*. [Online] 448 (7151), 313–317. Available from: doi:10.1038/nature05934.

Oliveros, J.C. (2007) *-Venny-. Venn Diagrams for comparing lists. By Juan Carlos Oliveros.* [Online]. 2007. Available from: https://bioinfogp.cnb.csic.es/tools/venny_old/venny.php [Accessed: 22 July 2019].

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., et al. (2011) Induction of human neuronal cells by defined transcription factors. *Nature*. [Online] 476 (7359), 220–223. Available from: doi:10.1038/nature10202.

Paonessa, F., Latifi, S., Scarongella, H., Cesca, F., et al. (2013) Specificity protein 1 (Sp1)-dependent activation of the synapsin I gene (SYN1) is modulated by RE1-silencing transcription factor (REST) and 5'-cytosine-phosphoguanine (CpG) methylation. *The Journal of biological chemistry*. [Online] 288 (5), 3227–3239. Available from: doi:10.1074/jbc.M112.399782.

Patzke, C., Han, Y., Covy, J., Yi, F., et al. (2015) Analysis of conditional heterozygous STXBP1 mutations in human neurons. *The Journal of clinical investigation*. [Online] 125 (9), 3560–3571. Available from: doi:10.1172/JCI78612.

Pereira, M., Pfisterer, U., Rylander, D., Torper, O., et al. (2014) Highly efficient

generation of induced neurons from human fibroblasts that survive transplantation into the adult rat brain. *Scientific Reports*. [Online] 4 (1), 6330. Available from: doi:10.1038/srep06330.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., et al. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*. [Online] 10 (10), 973–976. Available from: doi:10.1038/nmeth.2600.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., et al. (2011a) Direct conversion of human fibroblasts to dopaminergic neurons. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 108 (25), 10343–10348. Available from: doi:10.1073/pnas.1105135108.

Pfisterer, U., Wood, J., Nihlberg, K., Hallgren, O., et al. (2011b) Efficient induction of functional neurons from adult human fibroblasts. *Cell Cycle*. [Online] 10 (19), 3311–3316. Available from: doi:10.4161/cc.10.19.17584.

Qiang, L., Fujita, R., Yamashita, T., Angulo, S., et al. (n.d.) *Directed Conversion of Alzheimer's Disease Patient Skin Fibroblasts into Functional Neurons*. [Online] Available from: doi:10.1016/j.cell.2011.07.007.

Roma, G., Sardiello, M., Cobellis, G., Cruz, P., et al. (2007) The UniTrap resource: tools for the biologist enabling optimized use of gene trap clones. *Nucleic Acids Research*. [Online] 36 (Database), D741–D746. Available from: doi:10.1093/nar/gkm825.

Ruetz, T., Pfisterer, U., Di Stefano, B., Ashmore, J., et al. (2017) Constitutively Active SMAD2/3 Are Broad-Scope Potentiators of Transcription-Factor-Mediated Cellular Reprogramming. *Cell Stem Cell*. [Online] 21 (6), 791-805.e9. Available from: doi:10.1016/j.stem.2017.10.013.

Saitsu, H., Kato, M., Mizuguchi, T., Hamada, K., et al. (2008) De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nature Genetics*. [Online] 40 (6), 782–788. Available from: doi:10.1038/ng.150.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., et al. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research*. [Online] 39 (21), 9275–9282. Available from: doi:10.1093/nar/gkr606.

Schneuwly, S., Klemenz, R. & Gehring, W.J. (1987) Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. *Nature*. [Online] 325 (6107), 816–818. Available from: doi:10.1038/325816a0.

Schulze, K.L., Littleton, J.T., Salzberg, A., Halachmi, N., et al. (1994) Rop, a drosophila homolog of yeast Sec1 and vertebrate n-Sect/Munc-18 proteins, is a negative regulator of neurotransmitter release in vivo. *Neuron*. [Online] 13 (5), 1099–1108. Available from: doi:10.1016/0896-6273(94)90048-5.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*. [Online] 343 (6166), 84–87. Available from: doi:10.1126/science.1247005.

Sharma, S., Wongpalee, S.P., Vashisht, A., Wohlschlegel, J.A., et al. (2014) Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes & development*. [Online] 28 (22), 2518–2531. Available from: doi:10.1101/gad.248625.114.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., et al. (2008) Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition Margaret A Goodell (ed.). *PLoS Biology*. [Online] 6 (10), e253. Available from:

doi:10.1371/journal.pbio.0060253.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., et al. (2015) Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research*. [Online] 43 (18), e118. Available from: doi:10.1093/nar/gkv575.

Smith, D.K., Yang, J., Liu, M.-L. & Zhang, C.-L. (2016) Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports*. [Online] 7 (5), 955–969. Available from: doi:10.1016/J.STEMCR.2016.09.013.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., et al. (2011) Conversion of Mouse and Human Fibroblasts into Functional Spinal Motor Neurons. *Cell Stem Cell*. [Online] 9 (3), 205–218. Available from: doi:10.1016/J.STEM.2011.07.014.

Studer, L., Vera, E. & Cornacchia, D. (2015) Programming and Reprogramming Cellular Age in the Era of Induced Pluripotency. *Cell Stem Cell*. [Online] 16 (6), 591–600. Available from: doi:10.1016/J.STEM.2015.05.004.

Swanson, D.A., Steel, J.M. & Valle, D. (1998) Identification and Characterization of the Human Ortholog of Rat STXBP1, a Protein Implicated in Vesicle Trafficking and Neurotransmitter Release. *Genomics*. [Online] 48 (3), 373–376. Available from: doi:10.1006/GENO.1997.5202.

Takahashi, K. (2012) Cellular reprogramming--lowering gravity on Waddington's epigenetic landscape. *Journal of cell science*. [Online] 125 (Pt 11), 2553–2560. Available from: doi:10.1242/jcs.084822.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., et al. (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*. [Online] 131 (5),

861–872. Available from: doi:10.1016/j.cell.2007.11.019.

Takahashi, K. & Yamanaka, S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. [Online] 126 (4), 663–676. Available from: doi:10.1016/j.cell.2006.07.024.

Tanabe, K., Ang, C.E., Chanda, S., Olmos, V.H., et al. (2018) Conversion of human adult peripheral blood T cells into neurons. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 115 (25), 6470–6475. Available from: doi:10.1073/pnas.1720273115.

Tanackovic, G. & Krämer, A. (2005) Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Molecular biology of the cell*. [Online] 16 (3), 1366–1377. Available from: doi:10.1091/mbc.e04-11-1034.

Tang, Y., Liu, M.-L., Zang, T. & Zhang, C.-L. (2017) Direct Reprogramming Rather than iPSC-Based Reprogramming Maintains Aging Hallmarks in Human Motor Neurons. *Frontiers in molecular neuroscience*. [Online] 10, 359. Available from: doi:10.3389/fnmol.2017.00359.

Taniwaki, T., Haruna, K., Nakamura, H., Sekimoto, T., et al. (2005) Characterization of an exchangeable gene trap using pU-17 carrying a stop codon-betageo cassette. *Development, Growth and Differentiation*. [Online] 47 (3), 163–172. Available from: doi:10.1111/j.1440-169X.2005.00792.x.

Thiel, G., Greengard, P. & Sudhoft, T.C. (1991) Characterization of tissue-specific transcription by the human synapsin I gene promoter. *Proc. Nati. Acad. Sci. USA*.88.

Tian, J., Liu, Y., Zhu, B., Tian, Y., et al. (2015) SF3A1 and pancreatic cancer: new evidence for the association of the spliceosome and cancer. *Oncotarget*. [Online] 6

(35), 37750–37757. Available from: doi:10.18632/oncotarget.5647.

Toonen, R.F.G. & Verhage, M. (2007) Munc18-1 in secretion: lonely Munc joins SNARE team and takes control. *Trends in Neurosciences*. [Online] 30 (11), 564–572. Available from: doi:10.1016/J.TINS.2007.08.008.

Torper, O., Pfisterer, U., Wolf, D.A., Pereira, M., et al. (2013) Generation of induced neurons via direct conversion in vivo. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 110 (17), 7038–7043. Available from: doi:10.1073/pnas.1303829110.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., et al. (2016) Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature*. [Online] 534 (7607), 391–395. Available from: doi:10.1038/nature18323.

Tsunemoto, R., Lee, S., Szűcs, A., Chubukov, P., et al. (2018) Diverse reprogramming codes for neuronal identity. *Nature*. [Online] 557 (7705), 375–380. Available from: doi:10.1038/s41586-018-0103-5.

Tucker, K.L., Meyer, M. & Barde, Y.-A. (2001) Neurotrophins are required for nerve growth during development. *Nature Neuroscience*. [Online] 4 (1), 29–37. Available from: doi:10.1038/82868.

Turaç, G., Hindley, C.J., Thomas, R., Davis, J.A., et al. (2013) Combined Flow Cytometric Analysis of Surface and Intracellular Antigens Reveals Surface Molecule Markers of Human Neuropoiesis Eva Mezey (ed.). *PLoS ONE*. [Online] 8 (6), e68519. Available from: doi:10.1371/journal.pone.0068519.

Tycko, J., Myer, V.E. & Hsu, P.D. (2016) Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell*. [Online] 63 (3), 355–370. Available from:

doi:10.1016/J.MOLCEL.2016.07.004.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., et al. (2016) A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell reports*. [Online] 17 (4), 1193–1205. Available from: doi:10.1016/j.celrep.2016.09.079.

Uhlén, M., Björling, E., Agaton, C., Szigyarto, C.A.-K., et al. (2005) A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Molecular & Cellular Proteomics*. [Online] 4 (12), 1920–1932. Available from: doi:10.1074/MCP.M500279-MCP200.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., et al. (2015) Tissue-based map of the human proteome. *Science*. [Online] 347 (6220), 1260419–1260419. Available from: doi:10.1126/science.1260419.

Urigüen, L., Gil-Pisa, I., Munarriz-Cuezva, E., Berrocoso, E., et al. (2013) Behavioral, neurochemical and morphological changes induced by the overexpression of munc18-1a in brain of mice: relevance to schizophrenia. *Translational Psychiatry*. [Online] 3 (1), e221–e221. Available from: doi:10.1038/tp.2012.149.

Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., et al. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. [Online] 435 (7042), 646–651. Available from: doi:10.1038/nature03556.

Vera, E., Bosco, N. & Studer, L. (2016) Generating Late-Onset Human iPSC-Based Disease Models by Inducing Neuronal Age-Related Phenotypes through Telomerase Manipulation. *Cell reports*. [Online] 17 (4), 1184–1192. Available from: doi:10.1016/j.celrep.2016.09.062.

Verhage, M., Maia, A.S., Plomp, J.J., Brussaard, A.B., et al. (2000) Synaptic assembly of the brain in the absence of neurotransmitter secretion. *Science (New York, N.Y.)*. [Online] 287 (5454), 864–869. Available from: doi:10.1126/science.287.5454.864.

Victor, M.B., Richner, M., Olsen, H.E., Lee, S.W., et al. (2018) Striatal neurons directly converted from Huntington's disease patient fibroblasts recapitulate age-associated disease phenotypes. *Nature neuroscience*. [Online] 21 (3), 341–352. Available from: doi:10.1038/s41593-018-0075-7.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., et al. (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*. [Online] 463 (7284), 1035–1041. Available from: doi:10.1038/nature08797.

Waddington, C. (1957) *The Strategy of the Genes*. London: Allen and Unwin.

Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K.R., et al. (1998) Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature*. [Online] 394 (6691), 369–374. Available from: doi:10.1038/28615.

Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. (2014) Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*. [Online] 343 (6166), 80–84. Available from: doi:10.1126/science.1246981.

Wapinski, O.L., Lee, Q.Y., Chen, A.C., Li, R., et al. (2017) Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell reports*. [Online] 20 (13), 3236–3247. Available from: doi:10.1016/j.celrep.2017.09.011.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., et al. (2013) Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. *Cell*. [Online] 155 (3), 621–635. Available from: doi:10.1016/J.CELL.2013.09.028.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., et al. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*. [Online] 448 (7151), 318–324. Available from: doi:10.1038/nature05944.

Wernig, M., Tucker, K.L., Gornik, V., Schneiders, A., et al. (2002) Tau EGFP embryonic stem cells: An efficient tool for neuronal lineage selection and transplantation. *Journal of Neuroscience Research*. [Online] 69 (6), 918–924. Available from: doi:10.1002/jnr.10395.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., et al. (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature*. [Online] 385 (6619), 810–813. Available from: doi:10.1038/385810a0.

Wu, M.N., Littleton, J.T., Bhat, M.A., Prokop, A., et al. (1998) ROP, the Drosophila Sec1 homolog, interacts with syntaxin and regulates neurotransmitter release in a dosage-dependent manner. *The EMBO Journal*.17 (1).

Xie, H., Ye, M., Feng, R. & Graf, T. (2004) Stepwise reprogramming of B cells into macrophages. *Cell*. [Online] 117 (5), 663–676. Available from: doi:10.1016/s0092-8674(04)00419-2.

Yang, N., Ng, Y.H., Pang, Z.P., Südhof, T.C., et al. (2011) Induced Neuronal Cells: How to Make and Define a Neuron. *Cell Stem Cell*. [Online] 9 (6), 517–525. Available from: doi:10.1016/J.STEM.2011.11.015.

Yang, Y., Jiao, J., Gao, R., Le, R., et al. (2015) Enhanced Rejuvenation in Induced Pluripotent Stem Cell-Derived Neurons Compared with Directly Converted Neurons from an Aged Mouse. *Stem Cells and Development*. [Online] 24 (23), 2767–2777. Available from: doi:10.1089/scd.2015.0137.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., et al. (2013) Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. *Neuron*. [Online] 78 (5), 785–798. Available from: doi:10.1016/J.NEURON.2013.05.029.

Zhou, D., Zhang, Z., He, L.-M., Du, J., et al. (2014a) Conversion of fibroblasts to neural cells by p53 depletion. *Cell reports*. [Online] 9 (6), 2034–2042. Available from: doi:10.1016/j.celrep.2014.11.040.

Zhou, Y., Alimohamadi, S., Wang, L., Liu, Z., et al. (2018) A Loss of Function Screen of Epigenetic Modifiers and Splicing Factors during Early Stage of Cardiac Reprogramming. *Stem cells international*. [Online] 2018, 3814747. Available from: doi:10.1155/2018/3814747.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., et al. (2014b) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*. [Online] 509 (7501), 487–491. Available from: doi:10.1038/nature13166.

Appendix 1 Library gene list

1110004E09Rik	Abi3bp	Akr7a5	Armc9
1190002N15Rik	Abl2	Alas1	Armcx1
1700037H04Rik	Abr	Aldh7a1	Arpc3
1810041L15Rik	Abtb1	Aldoart1	Arrdc1
1810043G02Rik	Acaa2	Alkbh1	Asb1
2010111I01Rik	Acat3	Amigo1	Asb6
2210016L21Rik	Acer3	Ampd3	Asb8
2310003H01Rik	Acly	Amph	Ascl1
2310061I04Rik	Acot10	Angptl6	Asf1a
2410131K14Rik	Acot2	Ank	Asl
2510009E07Rik	Acot9	Ankrd13d	Asns
2610002J02Rik	Acta2	Ankrd27	Aspn
2610034B18Rik	Actl6a	Ankrd28	Asrgl1
2610301B20Rik	Adam17	Anks3	Atf3
2700081O15Rik	Adam23	Ano10	Atg12
3110043O21Rik	Adrbk1	Anxa1	Atg2b
8430419L09Rik	Aftph	Anxa3	Atg9a
8430427H17Rik	Aga	Anxa4	Atp13a2
9330159F19Rik	Agfg1	Anxa7	Atp6v0a1
9430020K01Rik	Agpat1	Ap2a2	Atp6v0c
A530054K11Rik	Agrn	Ap3s2	Atp6v1g2
A830010M20Rik	Agtpbp1	Apaf1	Atp8a1
Aacs	Ahi1	Apeh	Atp9a
Aagab	Ahnak	Aplp1	Atr
Aamdc	Ahr	Apod	Atrnl1
Aar2	AI413582	Apoe	AW209491
Aarsd1	AI593442	Arhgap24	AW551984
Abca1	Aig1	Arhgap28	Axl
Abcb8	Ajap1	Arid5b	B3galnt1
Abcf3	Ak1	Arl6ip5	B3gnt1
Abhd2	Akap12	Armc10	B4galt6
Abhd4	Akap5	Armc6	B9d1
Abhd6	Akap6	Armc8	Bag6

Bbs4
Bbs7
BC005764
Bcar1
Bcl2l11
Bcl7a
Bcl7b
Bcorl1
Bend4
Bend6
Bex2
Bgn
Bicd1
Birc5
Blvrb
Bmp4
Bmp7
Bscl2
Btbd2
Btbd7
Cabin1
Cadm1
Calm1
Calm2
Calm3
Calm4
Calm5
Calml3
Calml4
Camk2b
Camk4
Capg
Casc3
Casp12
Casp6
Casp7
Casp8
Cav1
Cby1
Ccdc181
Ccdc28b
Ccdc80
Ccdc97
Ccm2
Ccna2
Ccne2
Ccng2
Ccp110
Cd151
Cd200
Cd24a
Cd59a
Cd9
Cd99l2
Cdc23
Cdc25b
Cdc34
Cdh11
Cdipt
Cdk16
Cdk20
Cdkn1c
Cdkn2d
Cdon
Celf4
Cenpt
Cenpw
Cep170b
Cep19
Cep44
Cers4
Cers5
Cgnl1
Chac2
Chchd10
Chchd6
Chd7
Chek2
Chic1
Chpt1
Chrna4
Ciapin1
Cib3
Cic
Cisd3
Cited2
Ckb
Cks1b
Clasp2
Clcn3
Cltb
Cmas
Cndp2
Cnih2
Cnn2
Cnot10
Cnr1
Cnrip1
Cntn1
Coil
Col11a1
Col1a1
Col1a2
Col3a1
Col5a2
Copz2
Coq2
Coq3
Coro7
Cotl1
Cox19
Cox8b
Cplx2
Cpsf1
Cpt1c
Creb1
Creb3
Creb3l1
Creb3l2
Creb3l3
Creb3l4
Creb5
Crebbp
Crebl2
Crebrf
Crebzf
Creld2
Crlf1
Crtac1
Crtap
Csdc2
Csf1
Csgalnact2
Csnk1d
Cthrc1
Ctnna1
Ctps2
Ctsk
Ctsz
Cx3cl1
Cxcl12
Cxxc5
Cyba
Cyp51
Cyr61
D10Bwg1379e
D230025D16Rik
Daam1

Dab1
Dalrd3
Dap
Dbf4
Dbn1
Dbpht2
Dcaf5
Dcn
Dcx
Ddah1
Ddit3
Ddr2
Ddx25
Ddx39
Ddx41
Ddx55
Ddx56
Deaf1
Decr2
Dennd5b
Dffa
Dgkz
Dhcr7
Dis3l2
Dkc1
Dlc1
Dlx3
Dnajb2
Dnal1
Dner
Dnm1
Dnm2
Dock1
Dock4
Dohh
Dos
Dot1l
Dpp3
Dpp9
Dpy19l1
Dpysl4
Dpysl5
Dscam
Dse
Dsel
Dtd1
Dtx3
Dusp14
Dusp8
Dync1i1
Ebf2
Ebp
Ece1
Ecel1
Ecscr
Edem2
Edem3
Edil3
Eepd1
Efemp2
Eif2s3y
Eif4ebp1
Elavl2
Elf2
Elovl1
Elovl6
Elp3
Emb
Emc10
Emc9
Emd
Eml5
Emp1
Enah
Endod1
Eno1
Eno2
Epha5
Exd2
Exoc6
Exosc7
Extl2
F2r
Fads1
Fam114a1
Fam117a
Fam122a
Fam126a
Fam134b
Fam155a
Fam169a
Fam171b
Fam174a
Fam184a
Fam196a
Fam19a5
Fam21
Fam213b
Fam214b
Fam217b
Fam57b
Fam60a
Fam69a
Fam69b
Fam73a
Fas
Fbn1
Fbxl16
Fbxl2
Fbxl5
Fbxo30
Fbxo32
Fbxo42
Fbxo44
Fbxo46
Fcgrt
Fdft1
Fdps
Fech
Fez1
Fgf7
Fhl2
Fibin
Fkbp9
Flad1
Flna
Flot1
Flrt1
Flrt2
Flrt3
Fn1
Fn3k
Fosb
Fosl2
Foxn2
Foxp2
Frk
Fry
Fsd1l
Fstl1
G6pc3
Gaa
Gabarapl1
Gabpa
Gabrb3
Gadd45a
Gadd45g
Gale

Galnt9
Ganab
Gas1
Gas6
Gca
Gclc
Gdap1
Gfra1
Gga3
Ghr
Gins4
Gja1
Glce
Gli3
Glipr1
Glrb
Gm561
Gmds
Gmpr
Gnai1
Gnao1
Gnb4
Gng11
Gng2
Gon4l
Got1
Gpc3
Gpc4
Gpc6
Gpd2
Gpm6b
Gpnmb
Gpr125
Gpr137
Gpr137c
Gpr180
Gpr22
Gpr64
Gpr85
Gprasp1
Grb2
Grhpr
Gria2
Gria3
Grin1
Gse1
Gsn
Gsta4
Gtf2i
Gtpbp8
Gulp1
Gusb
H2-K1
Hadh
Haghl
Hap1
Haus3
Hdac6
Hdac9
Hddc3
Heatr3
Heca
Helz
Hepacam2
Hes6
Hexb
Hist1h1c
Hist1h2ba
Hist1h2bb
Hist1h2bj
Hist2h2bb
Hist3h2a
Hk2
Hmcn1
Hmga1
Hmga2
Hmgcl
Hmgxb4
Homer2
Hoxa7
Hoxd13
Hras1
Hsd17b7
Hspbp1
Htr3a
I830012O16Rik
Id3
Idi1
Ier2
Ier3
Iffo1
Ifi35
Ifit3
Ifitm2
Ifitm3
Ift88
Igf2
Igfbp7
Ikbip
Ikbkap
Ikbkg
Il11ra1
Il13ra1
Ilkap
Impdh2
Ina
Inhba
Ints3
Ip6k1
Ip6k2
Ipo13
Irf7
Itm2a
Jag1
Kat2a
Katnb1
Kctd17
Kdelc1
Kdelc2
Kdelr3
Kdm1b
Kif1a
Kif21a
Kif26b
Kif5a
Kif5c
Klf10
Klf4
Klhdc10
Klhl21
Klhl5
Krt33b
Lamb1
Lamc1
Lasp1
Lats2
Ldhb
Lgals3bp
Lgals8
Lhfp
Lhfpl2
Lhfpl4
Lhx3
Lin7a
Lipa
Lman2l
Lmbr1
Lmbrd2

Lmtk2
Lonrf2
Lox
Loxl1
Lpar6
Lpcat4
Lpin2
Lrrc17
Lrrc4c
Lrrtm4
Lsm10
Ltbp1
Lum
Mad2l1bp
Maf
Mafg
Magee1
Magi1
Maml2
Maoa
Map1b
Map2k7
Mapk10
Mapk8ip1
Mapk9
Mapre3
Mapt
March1
Marf1
Mblac2
Mbnl3
Mccc2
Mcm4
Mcm6
Mcm7
Mcts2
Mcu
Mdk
Mecr
Med22
Mef2c
Meis2
Men1
Mettl14
Mettl17
Mfap3
Mfap4
Mfge8
Mfn1
Mfng
Mgp
Mgrn1
Mgst1
Mkl2
Mkx
Mlec
Mllt11
Mmp14
Mn1
Moap1
Morn2
Morn4
Mpp1
Mpped2
Mrpl3
Mrps10
Msantd4
Msrb2
Msto1
Msx1
Mtap
Mtfp1
Mthfs
Mtmr7
Mum1l1
Mvk
Mxra8
Myh9
Myo10
Myod1
Myof
Myt1l
Nagk
Nalcn
Nap1l2
Nap1l3
Napb
Napg
Ncdn
Ndrg4
Ndufaf6
Nefl
Negr1
Neu1
Nexn
Nfat5
Nfatc1
Nfatc2
Nfatc2ip
Nfatc3
Nfatc4
Nfkb1
Nfkb2
Nfkbia
Nfkbib
Nfkbid
Nfkbie
Nfkbil1
Nfkbiz
Nfrkb
Nid1
Nid2
Nin
Nipsnap1
Nit1
Nkain1
Nlgn2
Nme7
Nmnat2
Nnmt
Nomo1
Nprl2
Nr2c2ap
Nrg1
Nrk
Nrn1
Nsdhl
Nsmf
Nt5dc2
Nubpl
Nucb1
Nudt10
Nudt22
Nup54
Nup85
Nup98
Nupr1
Nxn
Oat
Ocrl
Odf2l
Ogfod1
Ogfrl1
Olfm1
Olfml2b
Orai3
Os9
Osbpl1a

Osbpl2
Osgin2
Oxr1
Pacsin2
Pak3
Palm
Palmd
Pam
Pard6a
Parp8
Pbx3
Pcdh18
Pcdhb17
Pcdhb7
Pck2
Pclo
Pcolce
Pcp4l1
Pcsk2
Pcyt2
Pdcd2
Pdcd2l
Pddc1
Pde1c
Pde2a
Pde6d
Pdgfrl
Pdia4
Pdlim1
Pdpn
Pdxk
Peg3
Penk
Perp
Pfas
Pfkfb2
Pfkm
Pgam2
Pgap1
Pgd
Pgm2l1
Phgdh
Phldb2
Phtf1
Phtf2
Phyhipl
Pigq
Pigu
Pik3r3
Pim2
Pkd2
Pknox2
Pla2g4a
Plagl1
Plcb4
Plekha1
Plekhj1
Plin2
Plk2
Plxna3
Pmepa1
Pmp22
Pmvk
Pnck
Pnma2
Pnmal1
Pnmal2
Polr3e
Pom121
Postn
Pou3f2
Pou4f1
Ppa1
Ppap2b
Ppfia2
Ppic
Ppie
Ppil1
Ppm1d
Ppme1
Ppp1r16b
Ppp1r9a
Ppp3ca
Ppp3cb
Ppp3cc
Ppp3r1
Ppp3r2
Pptc7
Pqlc3
Prepl
Prim2
Prkcb
Prkcdbp
Prkcz
Prmt10
Prmt7
Prorsd1
Prpf40b
Prps1
Prune
Prune2
Psat1
Psmb10
Psmd13
Psmd9
Psmf1
Ptchd1
Ptgds
Ptgr1
Ptpn5
Ptrf
Pus7
Pwwp2b
Pxdn
Pygb
Qpct
R3hcc1
Rab36
Rab39b
Rab3c
Rab5c
Rabgap1l
Rabggta
Rabl5
Ralgds
Ralgps2
Ramp3
Ranbp6
Ranbp9
Rasal2
Rassf1
Rassf4
Rbbp8
Rbfox1
Rbm12
Rbms1
Rbmx2
Rbpj
Rcc2
Rcn1
Rcn3
Rdh11
Reep2
Reep3
Reps1
Rfc4
Rgmb
Rgs16

Rgs17
Rhog
Rhoj
Rimklb
Rims2
Ripk1
Rit2
Rnf121
Rnf122
Rnf145
Rnf157
Rnf167
Rnf170
Rnf182
Rnf19a
Rnf34
Rnft1
Rpp14
Rpp25l
Rpp30
Rpp38
Rprd1a
Rprm
Rps15a
Rrm2
Rsph3a
Rsph3b
Rspry1
Rtca
Rtn1
Rtp4
Rufy3
Rundc3b
Runx1
Ruvbl1
Ruvbl2
Rwdd2a
S100a1
S100a16
S100a6
S100pbp
S1pr2
Samd12
Sapcd1
Sbk1
Scaf8
Scamp5
Scg3
Scg5
Schip1
Scpep1
Scrn3
Sdc3
Sdc4
Sdr39u1
Sema5a
Sema6a
Sepp1
Sept5
Serinc5
Serpine2
Serpini1
Sestd1
Setd6
Sf3a1
Sfrp2
Sgce
Sgk1
Sgms1
Sgpp2
Sgsm3
Sh2d3c
Sh3d19
Sh3gl2
Shank1
Shc1
Shisa2
Shisa9
Shmt2
Siah2
Sigmar1
Sirt1
Sirt3
Sirt5
Six4
Slc22a17
Slc25a10
Slc25a14
Slc25a19
Slc25a20
Slc25a27
Slc25a44
Slc27a4
Slc29a1
Slc2a13
Slc2a6
Slc31a2
Slc33a1
Slc39a1
Slc39a7
Slit2
Slitrk5
Smad1
Smarca1
Smc4
Smim18
Smoc1
Smox
Smyd5
Snap23
Snap29
Snapc2
Snapc5
Snrnp40
Snurf
Snx10
Snx24
Sorbs3
Sos2
Sparc
Spata5
Spata7
Spats2l
Spcs3
Spock1
Spop
Spred2
Spryd3
Spryd4
Sqle
Srgap3
Srpk3
Srrm4
Srrt
Srsf12
Ssh2
Sspn
St18
St3gal4
St3gal5
St8sia2
St8sia4
Stk19
Stmn2
Stoml1
Stxbp1
Stxbp5
Sulf1

Sulf2
Sult2b1
Sumf2
Susd4
Svil
Syn1
Syn3
Syngr3
Synpo
Syp
Syt4
Tacc1
Taf1
Taf9b
Tagap
Tagap1
Tagln
Tagln2
Taok2
Tapbp
Tars2
Taz
Tbc1d8
Tbc1d9b
Tceanc2
Tcf3
Tcn2
Tdrkh
Tecpr1
Tfap2b
Tfcp2l1
Tfpi
Tgds
Tgfb1i1
Tgfbr2
Tgfbr3
Tiam1
Timp3
Tk2
Tle4
Tm4sf1
Tmcc3
Tmed5
Tmeff1
Tmem101
Tmem119
Tmem150a
Tmem176a
Tmem180
Tmem246
Tmem38a
Tmem39a
Tmem41a
Tmem47
Tmem55b
Tmem69
Tmem80
Tmem8b
Tmod2
Tmpo
Tnfaip8
Tnnt2
Tnnt3
Tnpo2
Tom1l2
Tomm40l
Tor1aip1
Tpgs1
Tppp3
Tpra1
Tprn
Tpst2
Tram1l1
Trim25
Trim3
Trmt10b
Trmt1l
Trnp1
Tro
Trp53
Trp53inp1
Tsc2
Tsfm
Tspan12
Tspan13
Tspan14
Tspan7
Tspyl4
Tspyl5
Tssc4
Ttc28
Ttc37
Ttl
Ttyh2
Tuba4a
Tubb3
Tubb4a
Tubb6
Tubg1
Tusc2
Txndc5
Tyms
Uap1l1
Ube2d1
Ube2l6
Ube2o
Ube2q1
Ube2ql1
Ubxn8
Uchl1
Umps
Unc119b
Unc80
Usp28
Usp32
Usp34
Usp37
Usp46
Utrn
Vamp2
Vamp8
Vash2
Vasp
Vcam1
Vgll3
Vim
Vmac
Vps13a
Vps45
Vstm2a
Wbp1
Wdr12
Wdr18
Wdr35
Wdr47
Wdr6
Wdr91
Wdr92
Wdyhv1
Wipi2
Wisp1
Wnt5a
Wwtr1
Xaf1
Ypel1
Zbtb1
Zbtb18
Zbtb6

Zbtb7a	Zfp35	Zfp763	Zkscan16
Zc3h10	Zfp36	Zfp781	Zkscan6
Zc3h12c	Zfp36l2	Zfp81	Zmiz2
Zcchc12	Zfp423	Zfp830	Zmym1
Zcchc18	Zfp444	Zfp874a	Zmym3
Zfand2b	Zfp449	Zfp933	Znrf3
Zfp259	Zfp503	Zfp938	Zswim1
Zfp266	Zfp512	Zfp952	
Zfp30	Zfp715	Zfyve27	